2014

EVOLUTION AND DIVERGENCE OF THE STRUCTURAL AND PHYSICAL PROPERTIES OF DNA BINDING BY METHYL-CYTOSINE BINDING DOMAIN FAMILY MEMBERS 2 AND 3

Jason Cramer
Virginia Commonwealth University

EVOLUTION AND DIVERGENCE OF THE STRUCTURAL AND PHYSICAL PROPERTIES OF DNA BINDING BY METHYL-CYTOSINE BINDING DOMAIN FAMILY MEMBERS 2 AND 3

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

by

Jason Matthew Cramer, B.S., M.Ed.
B.S., Virginia Commonwealth University, 1999
M.Ed., Virginia Commonwealth University, 2010

Advisor: David C. Williams Jr., MD, PhD
Associate Professor Department of Pathology and Laboratory Medicine
University of North Carolina- Chapel Hill
Chapel Hill, North Carolina

Virginia Commonwealth University
Richmond, Virginia
June, 2014

Acknowledgment

Survival of this experience would not have been possible without the unflagging support of numerous people. I am compelled to first acknowledge my family and friends. With their support, I mustered the courage to leave the teaching profession to pursue this degree. Throughout the process, friends and family metaphorically—and sometimes literally—kicked me in the butt so as to keep me moving forward. These people have truly been the most fantastic "pains in my butt" and I love them all!

I am deeply thankful to all current and past laboratory members of the Williams Lab who assisted in my scientific training and tolerated my annoyances, especially my friends Ninad Walavalkar and William Buchwald. Also, I would like to thank Neel Scarsdale, and members of the Matthew Hartman, Jessica Bell, William Barton, and Dr. Gordon Ginder labs for their assistance and tutelage.

Several other persons are deserving of my thanks, including members of my research committee: Dr. Carlos Escalante, Dr. Andrew Larner, Dr. Darrell Peterson, and Dr. Montserrat Samso. This group of scientists represents a tremendous source of inspiration, and their support and guidance of my research and personal development has been invaluable. I thank also Drs. Jan Chlebowski and Louis DeFelice for their support.

Last, but certainly not least, I would like to thank my advisor, Dr. David C. Williams, Jr., for his enduring guidance, patience, and understanding. Teaching new tricks to an "old dog" like myself has not been an easy task, but his mentorship endowed me with both tremendous scientific insight and the courage to finish this process despite the personal losses I have experienced in my time with him. For that, I am eternally grateful.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

1D	One-dimensional
^{1}H	Hydrogen nucleus (proton)
2D	Two-dimensional
3D	Three-dimensional
5mC	5-methyl-cytosine
5hmC	5-hydroxymethyl-cytosine
^{15}N	Nitrogen-15
Å	Angstrom
A.q.	*Amphimedon queenslandica*
AUC	Analytical ultracentrifugation
B_0	Magnetic field strength
BAH	bromo-adjacent homology domain
BER	Base excision repair
bp	Base pair(s)
BRCA1	Breast cancer associated-1
CC	Coiled-coil
CD	Circular dichroism
CGI	cytosine-guanine dinucleotide island
CHD	Chromodomain/helicase/DNA-binding domain
cMBD2	Chicken methyl-cytosine binding domain family member 2
CpA	Cytosine-phosphate-adenine dinucleotide
CpG	Cytosine-phosphate-guanine dinucleotide

CR	Conserved region
CR1	Conserved region 1 (coiled-coil region for p66α)
CR2	Conserved region 2 (zinc finger binding domain region for p66α)
D_2O	Deuterium oxide (deuterated water)
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
dsDNA	Double-stranded deoxyribonucleic acid
dsRNA	Double-stranded ribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylene-diamine-tetraacetic acid
E.m.	*Ephydatia muelleri*
FID	Free induction decay
FT	Fourier transformation
GC	Guanine-cytosine dinucleotide
HDAC	Histone deacetylase
hmCpG	Hydroxymethyl-cytosine-phosphate-guanine dinucleotide
H.s.	*Homo sapiens*
HSQC	Heteronuclear single quantum coherence
Hz	Hertz
IPAP	In-phase, anti-phase
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
K_a	Acid dissociation constant

K_D	Dissociation constant
kDa	KiloDalton
M9	Minimal media
MBD	Methyl-cytosine binding domain
MBD2a	Methyl-cytosine binding domain family member 2, isoform a
MBD2b	Methyl-cytosine binding domain family member 2, isoform b
MBD3	Methyl-cytosine binding domain family member 3
MBD3 L1	Methyl-cytosine binding domain family member 3-like 1
MBD3 L2	Methyl-cytosine binding domain family member 3-like 2
mCpG	Methyl-cytosine-phosphate-guanine dinucleotide
MeCP2	Methyl-cytosine binding protein-2
MHz	Megahertz
Mi-2	Chromodomain-helicase-DNA-binding protein Mi-2 homolog
MRE	Mean residual ellipticity
mRNA	Messenger ribonucleic acid
MTA	Metastasis associated protein
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser effect spectroscopy
NuRD	Nucleosome remodeling and deactylase
PAGE	Polyacrylamide gel electrophoresis
PCNA	Proliferating cell nuclear antigen
PDB	Protein data bank

Pf1	Filamentous phage particle 1
PHD	Plant homeodomain
pI	Isoelectric point
ppm	Parts per million
Q	Structure factor
ras	Rat sarcoma gene
RbAp	Retinoblastoma-associated protein
RBBP	Retinoblastoma binding protein
RDC	Residual dipolar coupling
rf	Radiofrequency
RING	Really interesting new gene finger domain
RMSD	Root mean square deviation
RNA	Ribonucleic acid
RNAi	Ribonucleic acid interference
σ	Chemical shift
S^2	Order parameter
SDS	Sodium dodecyl sulfate
siRNA	Small interfering ribonucleic acid
SPR	Surface plasmon resonance
SSR	Sum of squared residuals
SVD	Singular value decomposition
T	Tesla
TET	Ten-eleven translocation

TMS	Tetramethylsilane- $(CH_3)_4Si$
TpG	Thymine-phosphate-guanine dinucleotide
TROSY	Transverse relaxation optimized spectroscopy
UHRF1	Ubiquitin-like with PHD and RING finger domains protein 1
UV	Ultraviolet

Abstract

EVOLUTION AND DIVERGENCE OF THE STRUCTURAL AND PHYSICAL PROPERTIES OF DNA BINDING BY METHYL-CYTOSINE BINDING DOMAIN FAMILY MEMBERS 2 AND 3

By Jason Matthew Cramer

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

Virginia Commonwealth University, 2014

Major Director: David C. Williams, Jr., M.D., Ph.D., Associate Professor, Department of Pathology and Laboratory Medicine, University of North Carolina-Chapel Hill, Chapel Hill, North Carolina

The studies presented in this dissertation, *Evolution And Divergence Of The Structural And Physical Properties Of DNA Binding By Methyl-Cytosine Binding Domain Family Members 2 And 3,* pertain primarily to two key epigenetic regulators involved with the biological interpretation of methylated DNA marks. We provide insights into the emergence and evolution of the MBD2 and MBD3 and how those molecular entities influence heritable changes in gene activity. We further provide details regarding the mystery surrounding MBD3 function and the MBD2-mediated capacity of primitive animals to carry out methylation-specific epigenetic mechanisms. In chapter two, we describe the DNA binding properties of MBD2 and MBD3. This study provides information regarding previously unidentified MBD3 binding properties and potential biological function. In chapter three, we show that sponges demonstrate a MBD2-mediated capacity for binding methylated DNA sites, recruit NuRD components *in vitro*, and knockdown of MBD2 in the freshwater desmosponge, *Ephydatia muelleri*, promotes an abnormal growth phenotype.

Chapter 1- Introduction

When we refer to DNA, we often speak of the molecule as a "template" used in the production of proteins. While one may picture a machine assembly line wherein pieces are assembled to construct a whole, operational structure, DNA functionality often does not result in the production of identical parts. DNA serves as a manuscript that may be interpreted in a number of different ways with many different outcomes. This script may be altered, thereby changing meaning, and thus interpretation. Alteration may occur in the form of changes to letters and words or even sentence structure. But, modification may also involve no apparent change to the words, or their meaning. Subtle emphasis or de-emphasis of words, sentences, or whole sections of the script may occur leading to the disuse of those portions of script. When speaking of DNA, these subtle changes in emphasis often result in alteration of gene function.

Epigenetic phenomena are processes that cause heritable changes to gene expression without altering genetic sequences. The coding of these occurrences can reside in modifications to histones and DNA. Scientists study these events through a variety of biochemical, bioinformatics, and molecular biology approaches. Study of epigenetic phenomena along with the functional and mechanistic changes they cause has expanded not only our understanding of human health and disease, but also of the evolutionary processes driving vertebrate evolution.

Collected epigenetic information is translated into an understanding of the "epigenome". To better understand the biological significance of these phenomena, the epigenome must be mapped. Unfortunately, epigenetic programming is highly variable

and tends to be subject to the internal conditions of individual cells. Nevertheless, numerous research projects are underway with the aim of elucidating epigenetic signaling events on several levels.

This thesis represents one of those projects. Specifically, the information presented herein sheds further light on the physical mechanisms and dynamics of epigenetic machinery at both ends of the metazoan spectrum. By describing current protein-protein and protein DNA interactions in their evolutionary context we can better elucidate the processes by which new protein properties emerge, the underlying mechanisms by which biological systems function, and how the diversity of living forms in nature came to be.

1.1 DNA Methylation

DNA methylation is the most abundant and extensively characterized of the identified eukaryotic epigenetic modifications (1). From plants to mammals, DNA methyltransferases (DNMTs) carry out DNA methylation through the transfer of a methyl group from the methyl donor S-adenosyl methionine (SAM) to a cytosine (figure 1.1); the addition of a methyl group to the 5-position of the cytosine ring in a cytosine-guanine dinucleotide (CpG) forms 5-methylcytosine (5mC). In humans, this phenomenon is important for numerous processes including embryogenesis, stem cell differentiation, genomic imprinting, X chromosome inactivation, and has been associated with transcriptional silencing and tumorigenesis (1-4).

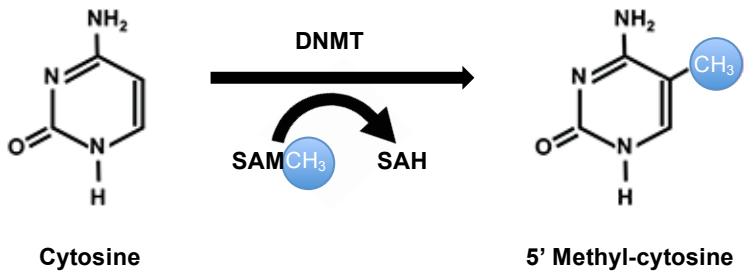


Figure 1.1: Image showing the process of DNA methylation.

1.1.1 DNA methylation is evolutionarily ancient

DNA methylation represents an important epigenetic signal found throughout the three domains of life, though the roles and targets of DNA methylation vary among the kingdoms of organisms. The process may have evolved in bacteria as a barrier against genome invasion by viral or other foreign DNA (3). The high degree of conservation in the catalytic motifs of DNMTs across prokaryotic and eukaryotic species also suggests that cytosine methylation is an ancestral mechanism.

In kingdom Eukaryota, 5mC and DNMTs are found in animal, plant, fungal, and protozoan lineages (5-7). Methylation patterns are also observed at the very base of the animal kingdom in sponges (8) presupposing the omnipresence of methylation and methylation machinery throughout higher levels of Eukaryota. Interestingly, this is not the case.

Patterns of DNA methylation changed during the transition from invertebrate to vertebrate life. Vertebrates tend to show global methylation often in promoter regions while invertebrates harbor “mosaic” methylation patterns in gene bodies (9). The global

methylation observed in vertebrates appears to correlate with an expansion of the DNMT and MBD families (10).

While 5mC and DNA methylation are abundant in vertebrates, these marks are not ubiquitous, and further, many invertebrates lack methylation altogether (9). Though several invertebrate organisms do not experience detectable methylation, including several commonly used invertebrate model organisms (e.g., yeast, fruit fly, and worm) (11), research continues to demonstrate the pervasiveness of cytosine methylation as an evolutionarily ancient genomic regulatory mechanism. Given that DNA methylation is a mechanism for regulating endogenous gene expression and reducing transcriptional noise, a reduction in unnecessary gene expression may have permitted the increase in gene number and in complexity that characterizes metazoans, i.e., animals (12).

1.1.2 DNA methylation in sponges

Sponges represent the most ancient and primitive metazoan lineage (13) and, despite their distinct morphological separation from other animals, studies of sponges have shed light on the molecular evolution of animals. Like higher animals, the sponge genome harbors genes crucial for growth, differentiation, cell-cycle control, cell adhesion, innate immunity, and allorecognition (14). Sponges also share in common with animals numerous transcription factor, sensory transduction, and cell adhesion genes (15). Sponges, however, have nearly two times more genes than eumetazoans (i.e., "true" animals) (16), discrediting the idea that animal evolution resulted from an increase in the number of genes.

Though the recently sequenced *Amphimedon queenslandica* genome indicates the numerous similarities in the molecular framework between sponges and other animals, much is still unknown regarding the epigenetic machinery in place at the dawn of metazoans. Thus, analysis of sponge epigenetic mechanisms and structures can potentially uncover the genomic innovations allowing for the multicellular life and shed light on the evolutionary underpinnings of cancer. Indeed, numerous "founder" genes associated with the appearance of metazoan multicellularity connect with cancer onset and progression (15).

Earlier analysis of sponge methylation levels detected up to 9.4% 5mC in a sample of 15 sponge species (8). The *A. queenslandica* genome exhibits germline cytosine methylation indicated by a depletion of CpGs with corresponding TpG and CpA dinucleotide excess relative to overall GC content. Four putative orthologs of DNMT1 were identified in *A. queenslandica* along with one copy of a DNMT-like gene (Pohlman D, unpublished data). Swiss-Prot analysis of *A.q.* revealed presence of partial sequences similar to the glycosylase domain of MBD4 along with other predicted MBD proteins, including a MBD2-like protein (8, 17).

Another sponge, *Ephydatia muelleri*, is an emerging model system to study aspects of animal development (14). *E. muelleri* harbors epigenetic machinery, discussed in detail later, that presumptively allows for the same epigenetic signaling as observed in higher animals. Two supposed orthologs of *DNMT1* were discovered in the *E. muelleri* genome along with one copy of a *DNMT3*-like gene. An ortholog of *MBD2* (i.e., *Mbd2/3*) was also detected (Pohlman D, unpublished data). Despite the presence of DNMTs and MBDs

in these two sponges, the methylation pattern (i.e., global or mosaic) has not yet been determined.

1.1.3 DNA methylation in mammals

DNA methylation occurs at 70-80% of CpGs in vertebrates although this dinucleotide represents a very small percentage of the base composition of DNA (18). The CpGs comprising the genome are primarily found concentrated into regions known as "CpG-islands" (CGIs) and represent a very small percentage (1-2%) of the genome in its entirety (19, 20). Approximately 60% of active genes contain CGIs in their promoter regions. These regions are largely unmethylated and less subject to CpG loss experienced by the rest of the genome whereas CpGs outside of CGIs tend to be very highly subject to methylation (19). Different cell types show different gene expression patterns, an observation that suggests that CpG methylation differs across different cell types.

DNA methylation patterns become distinctly altered early in mammalian embryonic development (21, 22). Fertilization initiates a period of genome-wide DNA demethylation and subsequent *de novo* methylation necessary for methylation reprogramming in somatic cells after implantation and resetting gamete-specific patterns in germ cells. Beyond these genome-wide changes, *de novo* methylation and demethylation occur during tissue-specific differentiation (23-25).

DNMT family members DNMT1, DNMT3A, and DNMT3B maintain mammalian DNA methylation patterns through the activity of a conserved C-terminal methyltransferase domain (5). DNMT1 preserves DNA methylation patterns during replication by adding methyl groups symmetrically opposed to methyl groups of hemi-

methylated CpG sites found on newly replicated DNA. DNMT1 localizes to these sites likely assisted by two proteins that bind hemi-methylated DNA: the cofactor ubiquitin-like, containing PHD and RING finger domains 1 (UHRF1) and proliferating cell nuclear antigen (PCNA) (21, 26). DNMT3A and DNMT3B have no preference for hemimethylated CpGs. Both establish methylation patterns during early stages of embryonic mammalian development through *de novo* DNA methylation activity (27).

Though cytosine methylation is considered to be a stable modification, a variety of mechanisms have been proposed for demethylation. During one such mechanism, 5mC is oxidized to form 5-hydroxymethylcytosine (5hmC) in mammals by the ten eleven translocation (TET) family of enzymes. Several mechanisms (e.g., the DNA base excision repair, BER, pathway) are proposed to occur subsequent to the conversion of 5mc to 5hmC and ultimately lead to a decrease of the methylation mark (28, 28, 29). These demethylation pathways may be important for normal developmental processes, as well as the global hypomethylation observed during cancer progression; downstream removal of 5hmC may be a critical event epigenetic gene regulation (28, 30, 31).

1.1.4 Aberrant DNA methylation and cancer

In normal cells, DNA methylation occurs primarily in repetitive genomic regions, including satellite DNA and parasitic elements (e.g., long and short interspersed transposable elements) (32). DNA methylation has been linked to regulation of gene expression and genome stability. CGIs around promoters are commonly unmethylated (33) with few exceptions (e.g., differentially methylated regions associated with gene imprinting). Despite early predictions gene body methylation has been associated with

transcription (34, 35). Methylation of gene promoter and enhancer regions, however, correlates with transcriptional silencing through interference of transcription factor binding and formation of heterochromatin by methyl DNA-binding proteins (MBDs) (36).

The relationship between disrupted DNA methylation patterns and cancer was first observed when hypomethylated cancer cells were identified and has since been extensively studied and argued (37, 37-39). Recent advances in our understanding of epigenetic modification has shown that human cancer cells undergo global changes in the epigenetic landscape including genome-wide hypomethylation concurrent with promoter CGI hypermethylation of tumor suppressor genes including BRCA1, VHL, and p16/INK4a (40); these alterations are now considered hallmarks of cancer. Further, these cells harbor numerous genetic alterations that, along with aberrant changes to the methylome, promote cancer progression.

The mechanism of gene silencing by promoter hypermethylation can occur by several modes including direct obstruction of transcriptional machinery and nucleosome remodeling activities that make transcriptional start sites inaccessible. DNA hypomethylation is often an early tumorigenic event that occurs in cancer cells primarily through demethylation in genomic repetitive regions (32, 40, 41). DNA demethylation can have mechanistic implications: removing methyl groups relaxes chromatin structure allowing histone acetylation and the binding of transcriptional complexes. Thus, loss of methylation can lead to activation of normally silent genes, including oncogenes such as those in the *ras* family (42). Relaxed chromatin also may be more prone to genetic mutation and chromosomal rearrangements (43).

1.2 Methyl-cytosine binding proteins 2 and 3, NuRD, and vertebrate evolution

The canonical MBD binds symmetrically related methyl-CpG dinucleotides. Though several MBD proteins have been identified, only MBD2 and MBD3 homologs have been identified in invertebrate genomes. MBD2/3 is the presumed invertebrate ancestor to the vertebrate MBD2 and MBD3 proteins; a single gene encodes MBD2/3, whereas two separate genes produce MBD2 and MBD3 (44). Vertebrate *Mbd2* and *Mbd3* possess nearly identical intron-exon structure and MBD2 and MBD3 exhibit 71% sequence homology. These and other observations have led to the supposition that a gene duplication event resulted in two genes distinct from Mbd2/3 at a time concurrent with the emergence of vertebrates (3).

Mammalian MBD2 selectively binds methylated DNA whereas MBD3 does not, although it does preferentially localize to methylated CpG dinucleotides (45). MBD3 differs from MBD2 at two amino acid residues critical for mCpG binding. In mammalian MBD3, a histidine residue (amino acid position 30) and a phenylalanine (position 34) replace lysine and tyrosine residues in MBD2, respectively. The latter substitution disrupts the interaction between the methyl group of methylcytosine and the MBD3 MBD; selectivity for methylated DNA is greatly reduced. Although these substitutions clearly abolish methyl-CpG selectivity, the biological purpose of MBD3 and much of its structure remain unclear.

MBD2 and MBD3 structural components primarily consist of a methyl-cytosine binding domain followed by a region implicated in binding the RbAp46/48 homolog p55 (46) and a C-terminal coiled-coil region. Human MBD2b shares 70% sequence homology with human MBD3 prompting *in vitro* comparisons between the two proteins. The human

MBD2a isoform possesses an N-terminal GR-rich region that likely serves as a DNA tether yet operates with lower affinity for methylated DNA in vitro than MBD2b (47).

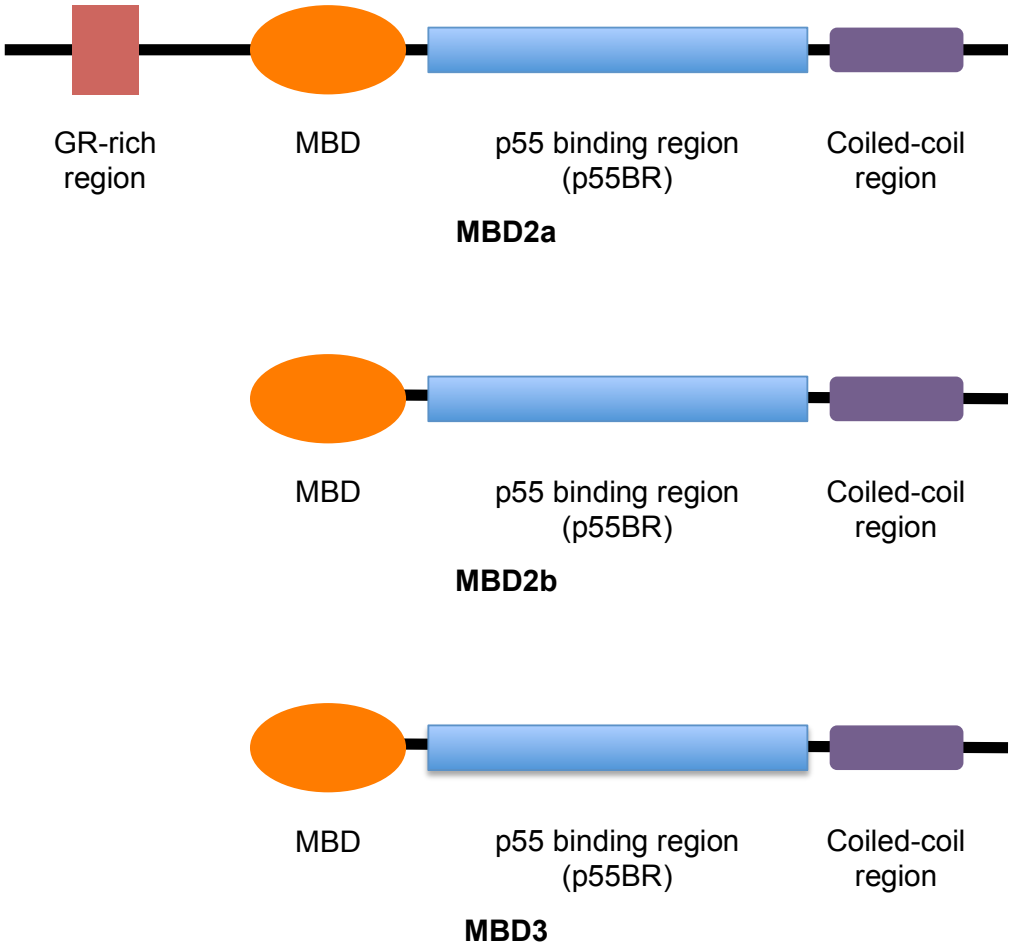


Figure 1.2: MBD2 and MBD3 domain organization. GR, glycine-arginine repeat region; MBD, methyl-binding domain.

Structural analysis of other MBD family members (i.e., MeCP2 and MBD1) elucidated the methyl-cytosine binding domain selectivity for mCpG dinucleotides: base-specific interactions between conserved arginine and tyrosine residues and the mCpG dinucleotide. A hydrogen bond network forms between arginine residues 22 and 44 and the symmetrically related guanine bases of the mCpG dinucleotide. The aliphatic portions of the arginine residues further pack against the methyl groups of the symmetric methyl-

cytosine bases. The critical tyrosine residue forms a hydrogen bond with the methyl group of a methyl-cytosine through water-mediated interaction (48, 49). Analysis of the MBD2 MBD solution structure elucidated a similar binding event, but further revealed a specificity for CGG sequences that enhances the selectivity of chicken MBD2 (cMBD2) for mCpGs driven in part by additional hydrogen bonding formed between lysine 30 and the second guanine of the mCGG sequence (50).

Comparison of cMBD2 and MBD3 MBD solution structures reveals high similarity between the two proteins and both complement the global fold of other methyl-cytosine binding domain proteins: the topology consists of a four-stranded β-sheet, a flanking α-helix, and a C-terminal loop (45, 50). The second and third beta sheets form a long, finger-like projection that extends into the major groove to form critical DNA-specific contacts. A long loop connects the strands forming this projection, and for cMBD2, residues found in this loop form contacts with DNA that further stabilize the interaction with methylated DNA (50). Notably, this loop is less well ordered for MBD3 (45); the absence of the amino acids crucial for methylation selectivity at the base of the projection destabilizes its overall structure, thereby disrupting the hydrogen bond network necessary for methylation-specific binding. Again, such information brings into question the functional role of MBD3.

Other observations further highlight the mystery surrounding mammalian MBD3 function: (1) Non-mammalian vertebrates have been identified that share high amino acid sequence identity with mammalian MBD3 yet preferentially bind methylated DNA, e.g., amphibian MBD3 (3, 47); (2) while *MBD2*-null mice experience only subtle defects *MBD3*-null mice die early in embryonic development (51); (3) both proteins associate

with the Mi-2/nucleosome remodeling and deacetylation complex (i.e., NuRD complex) in a mutually exclusive manner (52); (4) MBD3/NuRD bars reversion of murine stem cells to ground-state pluripotency (66); (5) MBD3/NuRD localizes to promoters (53, 54), gene bodies, and enhancers of active genes (54).

The NuRD complex is an ATP-dependent chromatin remodeling complex that facilitates multifaceted chromatin remodeling activities and is implicated in transcriptional silencing. The complex tethers to DNA through its associated MBD protein, thus coupling reading methylation marks with histone modification. As seen in figure 1.3, six core proteins comprise the complex and constitute its ability regulate gene expression at the chromatin level: MBD2 or MBD3; histone deactylase (HDAC) 1 or HDAC2; chromatin remodeling enzyme Mi-2α (CHD3) or Mi-2β (CHD4); p66α (GATA2A) or p66β (GATA2B); metastasis-associated (MTA) protein (i.e., MTA1, MTA2, or MTA3); retinoblastoma-binding protein (RBBP; also known as RbAp48) 4 or RBBP7 (RbAp46). The MBD protein targets the complex to DNA while HDAC1 or HDAC2 enzymatically removes acetyl groups from histone tails. The Mi-2 subunit carries out chromatin remodeling in part by repositioning nucleosomes. The p66α, p66β, RBBP4, and RBBP7 subunits are likely structural subunits of the complex that interact with histone tails. The p66α/β proteins contain two highly conserved regions: an N-terminal coiled-coil domain (CR1) and a C-terminal GATA-like zinc finger domain (CR2) (55, 56). The MTA protein is thought to localize the complex to different cell-specific targets by associating with transcription factors (57).

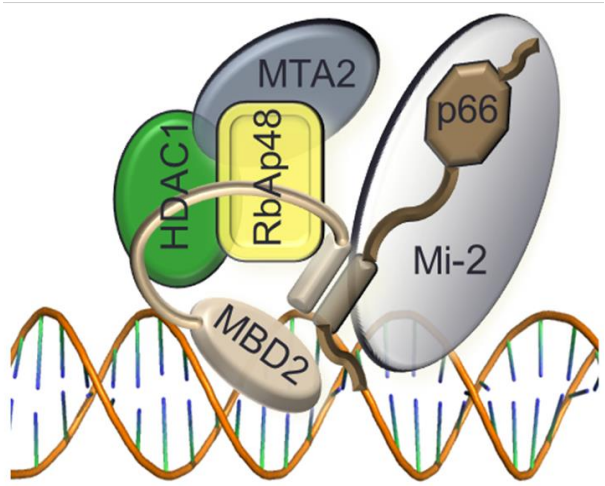


Figure 1.3: NuRD complex architecture. The human NuRD complex contains one homolog of six different core proteins: MBD2 or 3; HDAC 1 or 2; RbAp 46 or 48; p66α or β; Mi-2α or β.

The coiled-coil regions of MBD2 and MBD3 bind the p66α coiled coil domain (CR1) forming a high affinity heterodimeric complex that recruits Mi-2 and correlates with *in vivo* methylation dependent transcriptional repression (58). This interaction is unique amongst coiled coils in that the heterodimeric complex forms from anti-parallel association between two peptides that are largely monomeric in isolation. Helical content and specific electrostatic interactions between charged residues on the individual coiled coil domains drive the high affinity binding observed (59). This suggests that isoforms of MBD2, MBD3, and p66 exhibiting a reduction in helical content will bind with reduced affinity for p66. Indeed, this is the case for MBD3L1 and MBD3L2, homologues of MBD3 lacking a methyl-cytosine binding domain but containing a C-terminal coiled coil (59).

MBD2 and MBD3 contain a region implicated in binding the RbAp46/48 homolog p55 in Drosophila (46). The p55-binding region (p55BR) is found between the methyl-cytosine binding domain (MBD) and the coiled-coil domains of MBD2 and MBD3 and

recruit the NuRD core components RbAp48, HDAC2 and MTA2 (Desai, MA, unpublished data). This region has been determined to display a largely unstructured architecture in isolation and yet enhances the affinity of MBD2 to methylated DNA (Desai, MA, unpublished data).

Genes for NuRD complex components have been identified in the earliest metazoans including *Amphimedon queenslandica* and *Ephydatia muelleri* (unpublished data, Pohlman D.) but it is unclear if their products form a functional complex. We recently found, discussed in detail later, the *E. muelleri* MBD2 coiled-coil region binds the *E.m.* p66α coiled-coiled domain with low affinity. Further, the human MBD2 coiled coil domain binds *E.m.* p66aCR1 and vice versa with higher affinity than the two *E.m*. components in complex together (unpublished data). Surface plasmon resonance shows that the *Ephydatia* MBD2/3 methyl-cytosine binding domain binds methylated DNA with approximately 200-fold higher affinity than unmethylated DNA, similar to that of the human MBD2 (unpublished data). Isothermal titration calorimetry indicates that the coiled-coil region from the *E.m.* MBD2/3 binds to the human p66αCR1 with higher affinity than the weakly binding *Ephydatia* p66αCR1 isoform; the human isoforms bind each other with an affinity approximately 1,000-fold higher. These findings suggest that as coiled coil domains evolved in vertebrates, high affinity binding between these NuRD components developed so as to decrease competition with other coiled-coil proteins.

1.3 The coiled coil motif in evolution

Coiled-coils are an important structural motif involved in a diverse array of important protein interactions. They are a highly adaptable class of proteins both in terms of structure and function, and they are prevalent in all domains of life indicating the

evolutionary success of the motif. Coiled-coil domains are predicted in approximately 10% of eukaryotic proteins, and less than 5% in prokaryotes (60, 61). It has been suggested that the expansion of coiled coils and their interactions underlies metazoan cell and tissue structure variation (62).

The coiled-coil typically consists of two or more α-helices wrapped around one another. Coiled-coils are characterized by amino acid sequences forming a heptad repeat pattern. In this sequence, the first and fourth residues are hydrophobic, and residues in the fifth and seventh position are primarily charged or polar (63). Coiled-coil segments may be unfolded in monomeric form and then fold during interaction with a binding partner (64). A two-state transition often describes the folding and unfolding of coiled coils, i.e., unfolded peptide monomers cooperatively fold to form coiled-coil dimers or oligomers (63).

Despite the simplicity of the architecture, the coiled-coil motif exhibits considerable variation in terms of stability and dynamics. These characteristics may in turn determine the ability of coiled-coil domains to evolve; poorly packed and disordered proteins may show a high ability to evolve. Protein stability may then be tailored to the biological function. Coiled coils participate in numerous cellular functions and the simple yet dynamic nature of their structure may prompt binding with various partners and the formation of multiple complexes. However, binding partner selection appears to be an intricate process during which evolution may fine-tune each protein so as to enhance binding and allow for new biological functions, or the emergence of novel proteins (65).

Comparisons of sequence and structure provide insight regarding evolutionary relationships between proteins. The relatively simple repeating sequence of the coiled

coil motif suggests that it likely emerged multiple times during protein evolution. Small evolutionary changes in sequence and structure may result in large changes in binding dynamics, and thus biological function. The paths leading to novel functions likely proceed through states in which isoforms and relatively unstable structures emerge and coexist (65); one or more of these forms may perform new tasks beneficial for the organism resulting in the evolution of a new biological function. This scenario might very well be represented by the subtle changes over time in the coiled coil regions of p66α and MBD2 and MBD3.

The role of DNA methylation and methyl-cytosine binding domains has clearly changed during the course of metazoan evolution. The ancestral MBD may have played a role in gene regulation or modifying chromatin structure, but the expansion of the methylation landscape in vertebrate genomes created an environment in which new dynamic properties of regulatory proteins were required; expansion of the first MBD's functional role may have occurred concomitant to changes in the DNA methylome. Of course, it has been argued also that the methyl-CpG binding proteins provided both greater transcriptional control and protection against mutation, and thereby contributed to the global expansion of methylated DNA within the evolving vertebrate genome (3). Such contrasting arguments remain to be resolved. Thus, continued study of MBDs across the animal kingdom will provide further detail of the changing role of DNA methylation and gene regulation in animal evolution.

1.4 References

1. Smith ZD, Meissner A. DNA methylation: roles in mammalian development. *Nat Rev Genet.* 2013;14;3;204.
2. Hendrich B, Bird A. Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol Cell Biol.* 1998;18:6538.
3. Hendrich B, Tweedie S. The methyl-CpG binding domain and the evolving role of DNA methylation in animals. *Trends in Genetics*. 2003;19:269.
4. Ballestar E, Esteller M. Methyl-CpG-binding proteins in cancer: blaming the DNA methylation messenger. *Biochemistry and Cellular Biology*. 2005;83:374.
5. Suzuki MM, Bird A.
   DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet. 2008 Jun;9(6):. doi:* . 2008;9:465-76.
6. Colot V, Rossignol JL. Eukaryotic DNA methylation as an evolutionary device. *Bioessays*. 1999;21:402.
7. Ponger L, Li WH. Evolutionary diversification of DNA methyltransferases in eukaryotic genomes. *Molecular Biology and Evolution*. 2005;22:1119.
8. Regev A, Lamb MJ, Jablonka E. The Role of DNA Methylation in Invertebrates: Developmental Regulation or Genome Defense? *Molecular Biology and Evolution*. 1998;15:880.
9. Tweedie S, Charlton J, Clark V, Bird A. Methylation of genomes and genes at the invertebrate-vertebrate boundary. *Molecular and Cellular Biology*. 1997;17:1469.
10. Bird A, Wolffe AP. Methylation-induced repression--belts, braces, and chromatin. *Cell*. 1999;99:451.

11. Sarda S, Zeng J, Hunt BG, Yi SV. The evolution of invertebrate gene body methylation. *Molecular Biology and Evolultion*. 2012;29:1907.
12. Bird AP. Gene number, noise reduction and biological complexity. *Trends in Genetics*. 1995;11:94.
13. Coutinho CC, Fonseca RN, Mansure JJ, Borojevic R. Early steps in the evolution of multicellularity: deep structural and functional homologies among homeobox genes in sponges and higher metazoans. *Mechanisms of Development*. 2003;120:429.
14. Rivera A, Winters I, Rued A, et al. The evolution and function of the Pax/Six regulatory network in sponges. *Evolution and Development*. 2013;15:186.
15. Srivastava M et al. The Amphimedon queenslandica genome and the evolution of animal complexity. *Nature*. 2010;466:720.
16. Aury JM, et. al. Global trends of whole-genome duplications revealed by the ciliate Paramecium tetraurelia. *Nature. 2006 Nov 9;444(7116)*. 2006;444:171.
17. Albalat R, Martí-Solans J, Cañestro C. DNA methylation in amphioxus: from ancestral functions to new roles in vertebrates. *Briefings in Functional Genomics*. 2012;11:142.
18. Bird AP, Taggart MH. Variable patterns of total DNA and rDNA methylation in animals. *Nucleic Acids Res. 1980 Apr 11;8(7):1485-97*. 1980;8:1485-97.
19. Bird AP. CpG-rich islands and the function of DNA methylation. *Nature*. 1986;321:209-13.
20. Illingworth R, Kerr A, et. al. A novel CpG island set identifies tissue-specific methylation at developmental gene loci. *PLoS Biology*. 2008;6.

21. Auclair G, Weber M. Mechanisms of DNA methylation and demethylation in mammals. *Biochimie*. 2012;94:2202.

22. Cedar H, Bergman Y. Programming of DNA methylation patterns. *Annu Rev Biochem. 2012;81:97-117. doi:*. 2012;81:97-117.

23. Deng J, Shoemaker R, et. al. Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. *Nature Biotechnology*. 2009;27:353-60.

24. Kafri T, Ariel M, et. al. Developmental pattern of gene-specific DNA methylation in the mouse embryo and germ line. *Genes and Development*. 1992;6:705-14.

25. Monk M, Boubelik M, Lehnert S. Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development*. 1987;99:371-82.

26. Feng S, Jacobsen S, Reik W. Epigenetic reprogramming in plant and animal development. *Science*. 2010;330:622-7.

27. Okano M, Bell D, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*. 1999;99:247-57.

28. Wu SC, Zhang Y. Active DNA demethylation: many roads lead to Rome. *Nature Reviews Molecular Cellular Biology*. 2010;11:607-20.

29. Tahiliani M, Koh KP. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*. 2009;324:930-5.

30. Guibert S, Weber M. Functions of DNA methylation and hydroxymethylation in mammalian development. *Current Topics in Developmental Biology*. 2013;104:47-83.
31. Guibert S, Weber M. Chapter two - functions of DNA methylation and hydroxymethylation in mammalian development. In: *Current Topics in Developmental Biology.* Vol Volume 104. Academic Press:47-83.
32. Yoder JA, Walsh CP, Bestor TH. Cytosine methylation and the ecology of intragenomic parasites. *Trends in Genetics*. 1997;13:335-40.
33. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Reviews Genetics*. 2012;13:484-92.
34. Laurent. L., Wong E, et. al. Dynamic changes in the human methylome during differentiation. *Genome Research*. 2010;20:320-31.
35. Lister R, Pelizzola M. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*. 2009;462.
36. Kruczek I, Doerfler W. The unmethylated state of the promoter/leader and 5'-regions of integrated adenovirus genes correlates with gene expression. *The EMBO Journal*. 1982;1:409.
37. Feinberg AP, Tycko B. The history of cancer epigenetics. *Nature Reviews Cancer*. 2004;4:143-53.
38. Jones PA, Baylin SB. The epigenomics of cancer. *Cell*. 2007;128:683-92.
39. Schoofs T, Berdel WE, Müller-Tidow C. Origins of aberrant DNA methylation in acute myeloid leukemia. *Leukemia*. 2014;28:1-14.

40. Esteller M. Cancer Epigenetics for the 21st Century: What's Next? *Genes and Cancer*. 2011;2:604-6.

41. Widschwendter M, Jiang G. DNA hypomethylation and ovarian cancer biology. *Cancer Research*. 2004;64:4472-80.

42. Feinberg AP, Gehrke CW, Kuo KC, Ehrlich M. Reduced genomic 5-methylcytosine content in human colonic neoplasia. *Cancer Research*. 1988;48:1159-61.

43. Downs JA, Nussenzweig MC, Nussenzweig, A. Chromatin dynamics and the preservation of genetic information. *Nature*. 2007;447;951-58.

44. Hendrich B, Abbott C. Genomic structure and chromosomal mapping of the murine and human Mbd1, Mbd2, Mbd3, and Mbd4 genes. *Mammalian Genome*. 1999;10:906-12.

45. Cramer JM, Scarsdale JN. Probing the Dynamic Distribution of Bound States for Methylcytosine-binding Domains on DNA. *Journal of Biological Chemistry*. 2014;289:1294-302.

46. Marhold J, Brehm A, Kramer K. The Drosophila methyl-DNA binding protein MBD2/3 interacts with the NuRD complex via p55 and MI-2. *BMC Molecular Biology*. 2004;5:20.

47. Fraga MF, Ballestar E. The affinity of different MBD proteins for a specific methylated locus depends on their intrinsic binding properties. *Nucleic Acids Research*. 2003;31:1765-74.

48. Ohki I, Shimotake N. Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA. *Cell*. 2001;105:487-97.

49. Ho KL, McNae IW. MeCP2 binding to DNA depends on hydration at methyl-CpG. *Molecular Cell*. 2008;29:525-31.

50. Scarsdale JN, Webb HD, Ginder GD, Williams DCJ. Solution structure and dynamic analysis of chicken MBD2 methyl binding domain bound to a target-methylated DNA sequence. *Nucleic Acids Research*. 2011;39:6741-52.

51. Hendrich B, Guy J. Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. *Genes and Development*. 2001;15:710-23.

52. Le Guezennec X, Vermeulen M. MBD2/NuRD and MBD3/NuRD, two distinct complexes with different biochemical and functional properties. *Molecular and Cellular Biology*. 2006;26:843-51.

53. Günther K, Rust M. Differential roles for MBD2 and MBD3 at methylated CpG islands, active promoters and binding to exon sequences. *Nucleic Acids Research*. 2013;41:3010-21.

54. Shimbo T, Du Y, et. al. MBD3 Localizes at Promoters, Gene Bodies and Enhancers of Active Genes. *PLoS Genetics*. 2013;9.

55. Brackertz M, Gong Z, Leers J, Renkawitz R. p66alpha and p66beta of the Mi-2/NuRD complex mediate MBD2 and histone interaction. *Nucleic Acids Research*. 2006;34:397-406.

56. Feng Q, Cao R. Identification and functional characterization of the p66/p68 components of the MeCP1 complex. *Molecular and Cellular Biology*. 2002;22:536-46.

57. Lai AY, Wade PA. Cancer biology and NuRD: a multifaceted chromatin remodelling complex. *Nature Reviews Cancer*. 2011;11:588-96.

58. Gnanapragasam MN, Scarsdale JN, et. al. p66Alpha-MBD2 coiled-coil interaction and recruitment of Mi-2 are critical for globin gene silencing by the MBD2-NuRD complex. *Proc Natl Acad Sci U S A. 2011 May 3;108(18):7487-92. doi:*. 2011;108:787-92.

59. Walavalkar NM, Gordon N, Williams DC, Jr. Unique features of the anti-parallel, heterodimeric coiled-coil interaction between methyl-cytosine binding domain 2 (MBD2) homologues and GATA zinc finger domain containing 2A (GATAD2A/p66α). *Journal of Biological Chemistry*. 2013;288:3419-27.

60. Liu J, Rost B. Comparing structure and function between entire proteomes. *Protein Science*. 2001;10:1970-9.

61. Rose A, Schraegle SJ, Stahlberg EA, Meier I. Coiled-coil protein composition of 22 proteomes--differences and common themes in subcellular infrastructure and traffic control. *BMC Evolutionary Biology*. 2005;5:66.

62. Odgren PR, Harvie LW, Jr., Fey EG. Phylogenetic occurrence of coiled coil proteins: implications for tissue structure in metazoa via a coiled coil tissue matrix. *Proteins*. 1996;24:467-84.

63. Mason JM AK. Coiled coil domains: stability, specificity, and biological implications. *Chembiochem*. 2004;5:170-6.

64. Zitzewitz JA, Bilsel O, et. al. Probing the folding mechanism of a leucine zipper peptide by stopped-flow circular dichroism spectroscopy. *Biochemistry*. 1995;34:12812-9.

65. Gaspari Z, Nyitray L. Coiled coils as possible models of protein structure evolution. *BioMol Concepts*. 2011;2:199-210.

66. Rais, Y., Zviran, A., et al. Deterministic Direct Reprogramming of Somatic Cells to Pluripotency. *Nature*. 2013;502;7469;65-70.

Chapter 2- Probing the Dynamic Distribution of Bound States for Methyl-cytosine Binding Domains on DNA

2.1 ABSTRACT

Although highly homologous to other methylcytosine binding domain (MBD) proteins, MBD3 does not selectively bind methylated DNA and thus, the functional role of MBD3 remains in question. To explore the structural basis of its binding properties and potential function, we characterized the solution structure and binding distribution of the MBD3 MBD on hydroxymethylated, methylated and unmethylated DNA. The overall fold of this domain is very similar to other MBDs, yet a key loop involved in DNA binding is more disordered than previously observed. Specific recognition of methylated DNA constrains the structure of this loop and results in large chemical shift changes in NMR spectra. Based on these spectral changes, we show that MBD3 preferentially localizes to methylated and, to a lesser degree, unmethylated cytosine-guanosine dinucleotides (CpGs), yet does not distinguish between hydroxymethylated and unmethylated sites. Measuring residual dipolar couplings (RDCs) for the different bound states clearly shows that the MBD3 structure does not change between methylation specific and non-specific binding modes. Furthermore, RDCs measured for MBD3 bound to methylated DNA can be described by a linear combination of those for the methylation and non-specific binding modes, confirming the preferential localization to methylated sites. The highly homologous MBD2 protein shows similar but much stronger localization to methylated as well as unmethylated CpGs. Together, these data establish the structural basis for the relative distribution of MBD2 and MBD3 on genomic DNA and their observed occupancy at active and inactive CpG-rich promoters.

2.2 INTRODUCTION

The mammalian methylcytosine binding domain proteins (MeCP2 and MBD1-4) selectively bind symmetrically methylated CpGs through a common methylcytosine binding domain (MBD)(1) and likely arose from a gene duplication event of a single common ancestral protein (MBD2/3)(2). The preference of MBD2 for methylated DNA has been retained in both invertebrates and vertebrates; (3) however, the highly homologous MBD3 shows little to no preference for methylated DNA as a result of key differences in amino acids critical for DNA contact within the MBD (4, 5). Both proteins recruit a nucleosome remodeling and deacetylase (NuRD) complex (6) but in a mutually exclusive manner (7). The MBD2-NuRD complex has been specifically shown to promote methylation-dependent gene silencing and represents a potential therapeutic target for gene reactivation, (6, 8-13) while the function of the MBD3-NuRD complex has not been clearly delineated.

A recent study showed that MBD3 co-localizes with Tet1 and suggested preferential binding to hydroxymethylated CpGs (hmCpG) (14). Subsequent experiments, however, failed to show a binding affinity preference for hmCpG (15), but instead found that both MBD2 and MBD3 preferentially localize to transcriptional start sites with CGIs (16, 17). MBD2 predominates at methylated CGIs and the associated genes show reduced expression, while MBD3 appears to favor transcriptional start sites with unmethylated CGIs and is enriched at active promoters (16).

To help elucidate the structural differences between MBD2 and MBD3 and evaluate the recently proposed hydroxymethylation selectivity, (14) we determined the structure of MBD3 bound to DNA containing a single hmCpG dinucleotide. We observed that MBD3

adopts a very similar structure to that of MBD2. A critical loop connecting two anti-parallel β strands is less well defined in MBD3, but otherwise the two structures are nearly identical. Furthermore, we show that MBD3 does not specifically recognize hmCpG, but chemical shift analysis indicates that MBD3 binds differently to mCpG and spends a significant proportion of time on methylated sites. Occupancy depends on the number of unmethylated sites available and MBD3 demonstrates chemical shift averaging indicative of fast exchange between the methylated and non- specific binding modes.

Residual dipolar coupling (RDC) analysis confirms our findings by showing that MBD3 preferentially localizes to mCpG sites and that MBD3 adopts a very similar structure on mCpG, CpG, and hmCpG DNA. As expected MBD2 shows a strong preference for mCpG sites, exclusively localizing to the mCpG dinucleotide. We also find that MBD2 localization is influenced by unmethylated CpG density, and that MBD2 shows an unanticipated additional weak localization to hmCpG.

Taken together this information leads to a model in which the methylation specificity and occupancy of an MBD can be titrated by single amino acid substitutions. Importantly, the tendency to localize on a specific site does not necessarily translate into a global binding affinity preference. These data are consistent with recent studies showing that both MBD2 and MBD3 localize to transcription start sites associated with CGIs (16, 17, 31). The ability to condense chromatin and silence transcription at or near methylated CGIs reflects MBD2 high affinity and stable occupancy of mCpG sites. Hence, we propose that MBD3 evolved, at least in part, to counterbalance MBD2 on unmethylated CGIs. The presence of MBD3 at unmethylated CGIs could modify the

distribution of MBD2 and potentially mitigate strong silencing by the high- affinity, more strongly localizing MBD2 protein, thereby preserving bivalency with respect to transcription.

2.2.1 Use of NMR spectroscopy

Some of the key ideas of nuclear magnetic resonance (NMR) spectroscopy are discussed in this section. It should be noted that the description of NMR spectroscopy presented is not meant to offer a complete depiction of the process and the theory underlying its performance. The purpose of the material here is to provide some background that may allow the reader to better understand the NMR-derived data presented later in this chapter.

NMR spectroscopy represents a powerful tool for studying how biology works at the molecular level. The process provides access to information useful for studying several aspects of chemical biology on the atomic level, including: analysis of protein structure, protein folding dynamics, and protein-protein and protein-DNA interactions. Thus, NMR spectroscopy enhances our capacity to collect structural data from which atomic-level functional understanding can be developed.

2.2.1.1 Key principles

In quantum-mechanical theory, each atomic nucleus possesses a property known as *spin*. From a classical viewpoint, this attribute allows each nucleus to behave in a manner similar to a spinning top toy or a gyroscope. On a smooth surface a top spins freely about its main axis while also moving in a circular motion (i.e., *precession*) dictated by the

torque exerted by its weight. Figure 2.1, left image, represents this motion. In the case of the nucleus, the torque on the nuclear spin results from interaction with the magnetic field. The spinning top analogy is useful for explaining how the magnetic moments of the nuclei couple with an applied radiofrequency (*rf*) pulse and ultimately lead to a macroscopic oscillating magnetic field detected by a coil within the spectrometer.

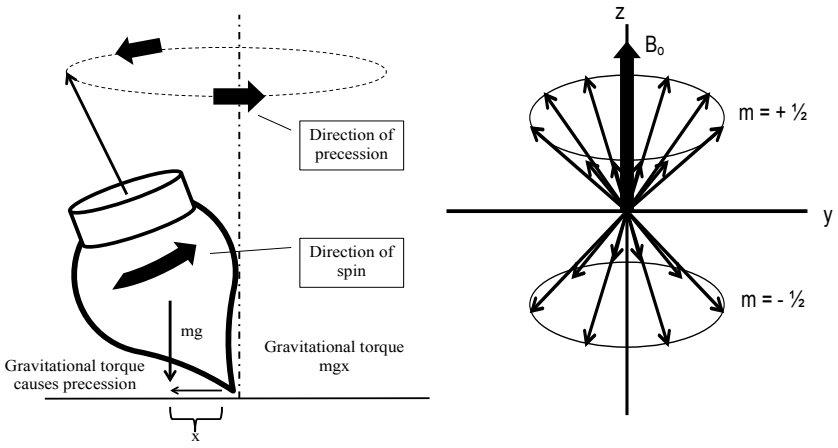


Figure 2.1: Nuclear spin. Left: a depiction of a spinning top toy. Right: a depiction of the two different allowed spin states of a ^{1}H nucleus.

In a sample of spin 1/2 nuclei in an external magnetic field, two populations will emerge, one aligned with the field, and the other against. Nuclei oriented against the direction of an external magnetic field (designated -1/2) associate with a higher energy spin state, and those oriented with the direction of the field (+1/2) associate with lower energy (figure 2.1, right; figure 2.2A). The energy difference between the two spin states is always very small but nevertheless dependent on the strength of the external magnetic field. That is, when the external field strength is zero, the two spin states are equally populated and have the same energy. When the field strength increases, the spin states diverge creating two distinct energy states. The ability to use strong magnets to

manipulate these populations of nuclei serves as the basis of NMR spectroscopy.

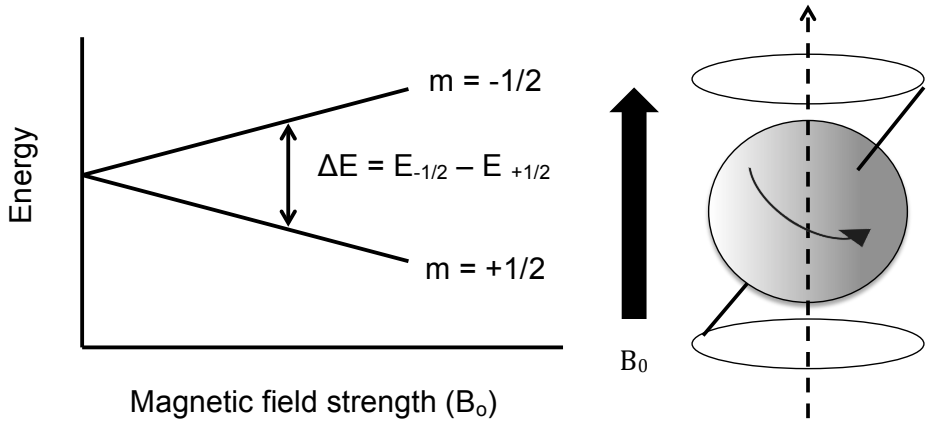


Figure 2.2: A) Energy versus magnetic field plot. B) Depiction of the magnetic dipole.

Even in a magnetic field of very high strength, the energy difference between two spin states is very small and is represented in MHz with the actual energy difference again depending on the strength of the field and the nucleus involved. For the nuclei used in this body of work, primarily hydrogen (^{1}H- proton) and nitrogen-15 (^{15}N) nuclei, the energy difference between their spin states depends upon their respective magnetic moments. Irradiation of the sample with *rf* energy that matches the energy difference between the two spin states may transition some of the nuclei in the lower energy state to the higher energy state, and vice-versa, at room temperature. Since excitation energy applied to the sample falls within the radiofrequency range, NMR spectroscopy represents a very mild manner by which molecular structure may be probed.

Although magnetic fields are uniformly applied to samples, the actual magnetic field experienced at the level of the nucleus depends upon several environmental factors;

scanning the energy absorption by nuclei provides highly useful information regarding the chemical environment in which nuclei exist in a sample. However, the absorption of radiation is very difficult to detect since the difference between the populations in the two spin states is very small, as discussed above. To heighten the ability to detect absorption of radiation, relatively high concentrations of the nuclei observed must be present.

Detection sensitivity depends on the characteristics of the nucleus observed, the abundance of the nucleus, and the energy difference in the two spin states. Thus, stronger magnets that produce larger magnetic fields are in continuous development. Unfortunately, carbon-12, which is highly abundant, produces no signal during NMR analysis. This is also the case for nitrogen-14. On the other hand, carbon-13 and nitrogen-15 isotopes will produce signal, hence their use in NMR studies is highly valuable.

Hydrogen nuclei are the most sensitive to NMR methods and the most extensive experimentation has been carried out using proton NMR. One might think that since all protons have the same magnetic moment, then all hydrogen nuclei will behave the same in a magnetic field. That is, hydrogen atoms might be expected to emit resonance signals at the same field versus frequency values. Fortunately, this is far from true. Protons in different compounds respond very differently in NMR experiments. The reason for the difference in behavior lies outside of the nucleus, in the electron cloud.

2.2.1.2 Shielding and Chemical shifts

As charged particles, electrons will respond to an external field by moving in a manner that generates a secondary magnetic field opposing the stronger external field. It has been shown that naked protons resonate at lower field strength than those hydrogen

nuclei involved in covalent bonds. This behavior results from the secondary field, generated by the movement of electrons, which *shields* the proton nucleus from the applied external field. Thus, the magnetic field strength must increase in order to bring about *resonance* (absorption of radiofrequency energy).

The ability to achieve excitement and resonance depends upon the degree of shielding exerted by surrounding electrons. The degree of shielding experienced by a proton is determined by the electronegativity of its covalent binding partner and of surrounding nuclei. For example, a hydrogen atom covalently bound to a bromine atom will be more *de*-shielded than one bound to a carbon atom, since the electronegativity difference between hydrogen and carbon is small. Further, a hydrogen atom in a chemical environment where it is both covalently bound to an atom with higher electronegativity and involved in an intermolecular interaction with a different atom of higher electronegativity (e.g. a hydrogen bond), will be further de-shielded; the proton nucleus will resonate at a higher frequency than that of a fully shielded nucleus.

The usefulness of NMR results from the interaction of nuclei with surrounding electrons and other nuclei in the molecule and sample solution. As mentioned earlier, high electron density around a nucleus produces a field that opposes the external field. The resonance frequency associated with this phenomenon scales with the magnetic field strength. Hence if we normalize by field strength, the chemical shift will be the same across spectrometers. Chemical shifts due to changes in chemical environments are taken into account by expressing the shift as a relative change in frequency with respect to the reference compound. To produce an unambiguous signal location, frequency differences (in Hz) between the reference signal and all other signals must be corrected by dividing

those differences by the frequency of the spectrometer (in MHz). Since this calculation itself will produce very small values, the quotient is multiplied by one million. The resulting value is known as the *chemical shift* (σ) and is reported in parts-per-million, ppm, a unit that reflects the normalization by the MHz field strength as described previously.

The chemical shift depends on the orientation of the sample to the magnetic field. Since the samples used in our studies are in solution, the rapid tumbling of the molecules leads to averaging of the chemical shifts thereby producing sharp resonance peaks. For large molecules and molecular complexes, the rate of tumbling is slower which broadens the resonance peaks observed and obscures the spectrum. Special conditions are met for solid or semi-solid environments so as to collect spectra, as will be presented later in the *residual dipolar coupling* discussion.

Again, high electron density increases shielding effect, such that the resonance frequency is lower. The result is an upfield shift and a decrease in the chemical shift towards the reference. Conversely, low electron density causes a downfield shift and an increase in σ. Although chemical shift data is highly useful for examining protein structure, chemical shift information is itself insufficient to determine protein structure.

The magnetic field range representing the resonances of protons in different chemical environments is very small (see figure 2.3). Detecting a ppm difference can be described as equivalent to detecting a one-millimeter difference in one-kilometer distances. Current application of NMR involves strong magnets that allow observation of proton resonances within most organic compounds between 0-14 ppm. Strong magnetic fields and high sensitivity may produce signals that are distinct and well separated, signal

detection and assignment in an unambiguous manner is very difficult without a numerical locator. To overcome this issue special reference compounds are added to samples so as to detect individual signals in a spectrum relative to the signal of the reference substance.

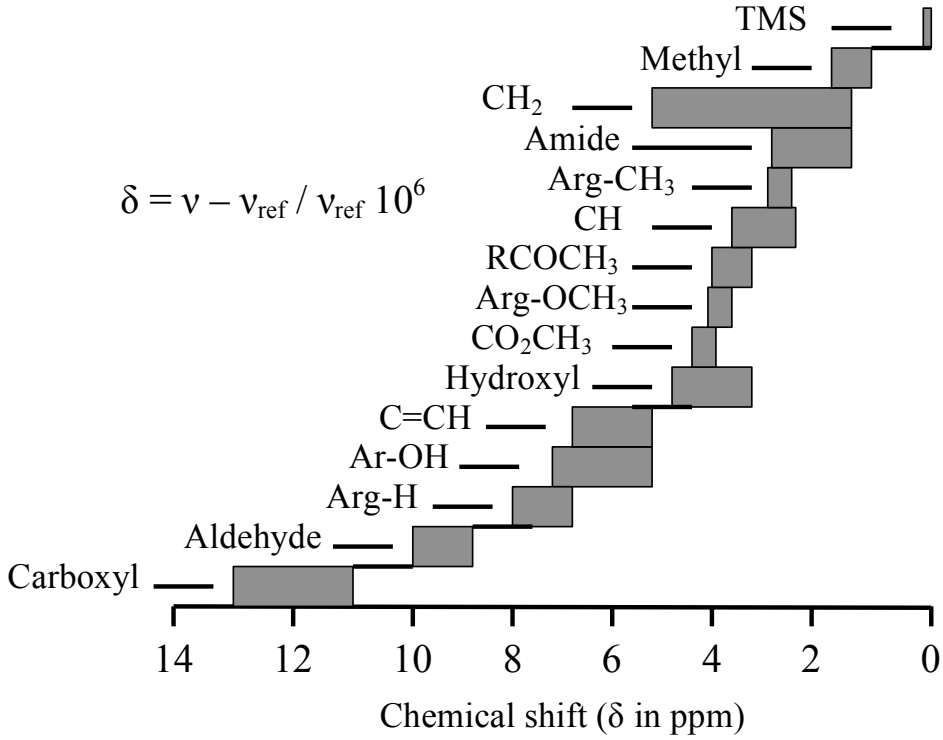


Figure 2.3: Magnetic field range representing the resonances of protons in different chemical environments.

The frequency dependence of the absorption can be transformed into a time dependence and vice-versa through Fourier transformation. In practice, a timed radiofrequency pulse is applied to a sample to excite the observed nuclei and then watch those nuclei return to their equilibrium population distributions in the magnetic field (aligned with or against the external field). During the return to equilibrium, known as the free induction decay (FID), an oscillating signal is produced which then decays over time

but not before a measurable oscillating current is generated in a coil of wire. The time domain data can be transformed into a frequency spectrum (figure 2.4) wherein the rate of decay determines the peak width and the oscillation determines the frequency of the peak. Multiple FIDS can be added to produce peaks with more signal-to-noise than a single FID.

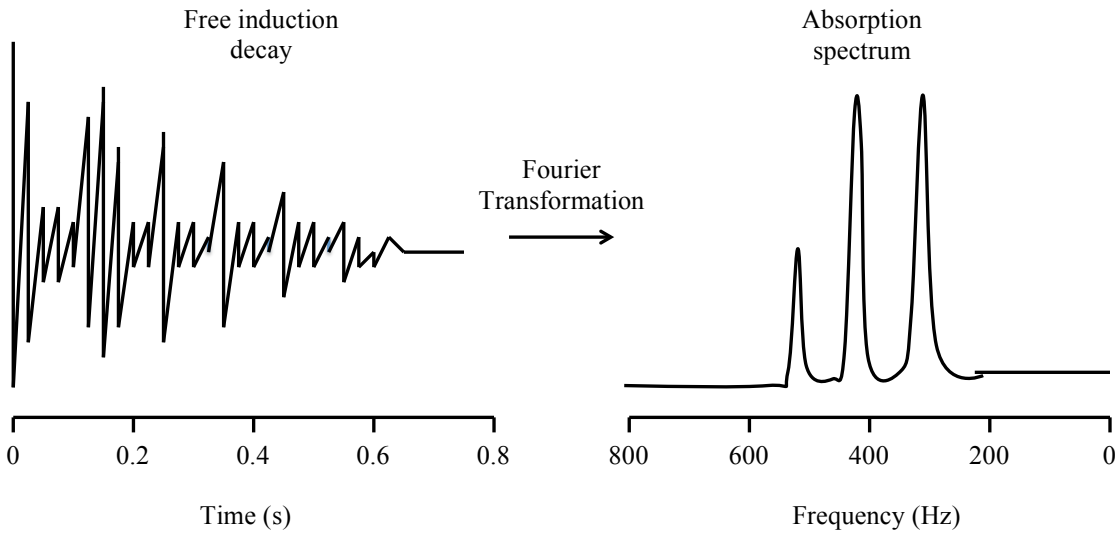


Figure 2.4: Free induction decay of nuclei in a magnetic field.

2.2.1.3 Relaxation times and chemical exchange

During an experiment a *rf* pulse orients nuclei in a specific direction, with precession about the direction of the external field. When the pulse ends, the nuclei in the sample will return to the equilibrium conditions and populations (spin states) associated with their environment. The rate at which each spin returns to its equilibrium conditions depends on interactions with its neighbors (surrounding nuclear spins). This rate also depends on fluctuations in the fields experienced as molecules tumble.

If a proton has a different chemical shift for two different states (represented by the

following simple expression: A↔B) then the proton resonates at two different frequencies. If the rate of exchange between the two states (*chemical exchange*) is very slow, then two distinct, sharp peaks will be observed in the NMR spectrum. Conversely, if the exchange occurs very rapidly, states A and B will interconvert repeatedly during the test, and the two frequencies will average, thus causing observance of a single sharp peak located at the weighted average between states A and B. This phenomenon indicates the fraction of time spent in state A versus B; if the weighted average is closer to A than B, then occupancy in state A is higher than that of state B and a peak will be observed closer to the peak associated with state A. During intermediate chemical exchange, the two lines observed in a slow exchange scenario will appear as if they had broadened and coalesced into a single broad peak. However, even in the fast exchange regime, if a large frequency difference exists between the two states, peak broadening may occur similar to that of intermediate exchange.

2.2.1.4 Two-dimensional (2D) NMR

The preceding information mostly pertained to acquisition of a spectrum for a specific nucleus (one-dimensional NMR- 1D NMR) by applying a frequency pulse and analyzing the frequencies associated with a FID. In order to collect more accurate structural information about molecules, utilizing multidimensional NMR becomes necessary, whereby "cross-peaks" linking two resonances are identified. For the purposes of the studies presented in this chapter, ^{1}H and ^{15}N resonances are analyzed, through space and through chemical bonds so as to identify the spatial relationship between those nuclei.

Whereas 1D NMR involves application of a single pulse, followed by FID and data collection, the multidimensional process begins with application of several radiofrequency pulses and time delays after which a FID is collected. The type of experiment determines the number and length of the pulses and delays. This experiment is repeatedly carried out and the signal collected is added. For the purpose of this explanation, the qualitative details of the pulses and delays will be excluded.

Ultimately, the collected data for the ^{1}H and ^{15}N resonances undergoes Fourier transformation and becomes represented in a two-dimensional plot where the frequencies for the two nuclei are plotted against one another. Nuclear spins of different non-bonded nuclei (usually ^{1}H-^{1}H) may be measured through space as a result of interactions (coherence transfer) between magnetic dipoles of the two nuclei. The through-space interaction of magnetic dipoles is known as the *Nuclear Overhauser Effect* (NOE) and analysis of the interactions is termed *Nuclear Overhauser Effect SpectroscopY* (NOESY). NOE interaction between two magnetic dipoles varies as the inverse sixth power of the distance between the two dipoles. Hence, only interacting nuclei very close to one another (within approximately 5Å) will give rise to measurable NOEs.

Another informative process involves detection only of pairs on covalently linked ^{1}H and ^{15}N nuclei and is known as *heteronuclear single quantum coherence* (HSQC). Each peak on an HSQC contour map represents a pair of covalently bound ^{1}H-^{15}N nuclei; by convention, the x-axis represents the proton resonance frequency and the y-axis represents the ^{15}N. HSQC spectra provide highly informative data for assigning observed resonances to specific amino acids in a protein structure, except proline, which has no backbone N-H. HSQCs can provide further detail of protein structure from additional

resonances provided by side-chains containing amides.

Structural determination from NMR data involves assigning resonances to specific nuclei within the structure via analysis of ^{1}H-^{15}N (or other) HSQCs and sequential ^{1}H-^{1}H NOEs. A three-dimensional model of the structure may be calculated through the use of numerous constraints provided by distance information from NOEs and dihedral angle information. The constraint information is then converted to a family of similar structures that satisfy the experimental constraints.

2.2.1.5 Residual dipolar coupling

Protein structure determination has traditionally relied on the ^{1}H-^{1}H NOE and various other constraints. Chemical shift analyses offer information pertaining to local constraints on dihedral bond angles, whereas NOE data can relate atoms of a protein far apart in the amino acid sequence, but close in space. Structural calculation often requires a large amount of NOE data, but nevertheless can produce precise structural detail. In certain situations, though, a lack of NOE cross-peaks reduces the ability to determine the relative orientation of protein domains (e.g., non-globular or multi-domain proteins, intermolecular complexes between proteins and between protein and DNA). Thus, a demand exists for techniques that provide both local and long-range structural information.

Interactions between pairs of magnetic dipoles in the presence of an external magnetic field are referred to as *dipolar couplings*. This type of interaction is similar in nature to a through-bond (scalar) interaction between nuclei, but unlike a scalar interaction, dipolar couplings average to zero through normal rotational tumbling of

molecules in solution. This phenomenon cannot be observed unless the rotational movement of molecules in solution becomes restricted to a preferred orientation with respect to the magnetic field. Introducing a weak degree of molecular alignment allows for measurement of the residual ^{1}H-^{15}N dipolar-coupling interaction whose magnitude depends on the degree of alignment and the angle formed between the internuclear vector and the magnetic field. Hence, in this situation these couplings are referred to as *residual dipolar couplings* (RDCs).

The information provided by RDCs directly relates to the distance between nuclei and their relative orientation regarding a molecular reference frame defined by the nature of the molecular alignment (see figure 2.5). Achievement of a weak degree of alignment is critical to the acquisition of this information since the dipolar coupling interaction is very strong; the effect can produce peaks wider than the spectrum normally collected. But if only a small degree of alignment is introduced, then this effect shows up as a very small chemical shift change (i.e., within a few Hertz) that compares to an otherwise very sharp spectrum from a protein tumbling freely in solution. Therefore, the nature of the alignment medium must allow for tuning, so that an appropriate degree of solute alignment can be achieved. The alignment medium must also be relatively stable in varying experimental conditions while minimizing adverse solute interactions that might prompt unfavorable alignment conditions (e.g., interactions that adversely affect rotational tumbling, thereby altering relaxation times).

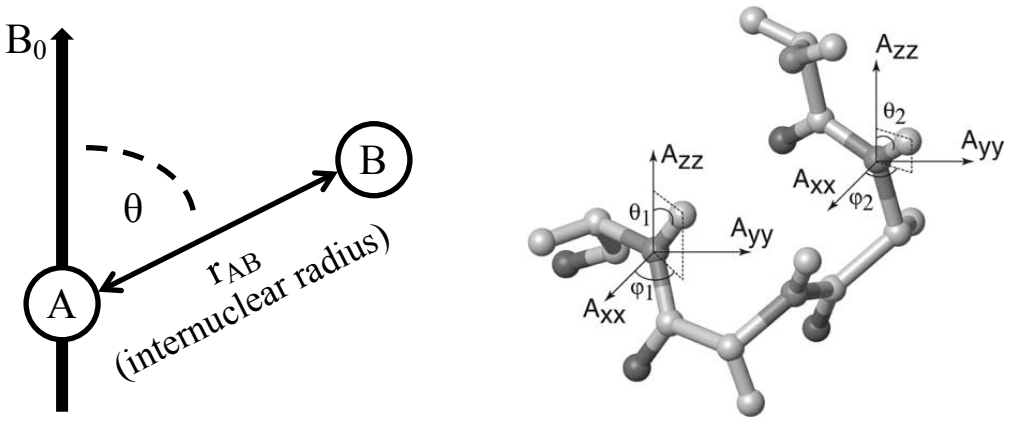


Figure 2.5: Relative orientation of nuclei in a magnetic field regarding a molecular reference frame defined by the nature of the molecular alignment.

Though several forms of alignment media are available, the filamentous bacteriophage particle, Pf1, was used as the alignment medium for the work described in this chapter. The bacteriophage particle will align in a preferred orientation in a magnetic field and influence the partial orientation of other solute molecules. Pf1 is highly negatively charged and can, therefore, electrostatically interact with molecules that have positively charged surface regions, as is the case with MBD2 and MBD3. For our studies, though, alignment was carried out in the presence of DNA; the DNA is very negatively charged and dominates the protein interaction. Hence, the repulsive interaction between the DNA and Pf1 drives the partial alignment. Tuning the concentration of phage to achieve sufficient alignment becomes necessary wherein the virus particles adopt a preferred orientation in the magnetic field. In this preferred orientation the particles point in the same direction with respect to the direction of the external magnetic field. Other molecules in solution will then adopt an orientation in response to that of the phage particles.

In practice, RDC analysis provides detail of the orientation of dipolar coupling

vectors, and thus the chemical bond orientation, in a protein that connect atoms experiencing homonuclear (^{1}H-^{1}H) and heteronuclear (^{1}H-^{15}N, CαHα, CαC', C'N) dipolar coupling interactions. For this body of work, coupling was measured between the amide ^{1}H and ^{15}N atoms so as to measure the angle between the amide bond and the alignment tensor, and thus, the relative orientations of NH bonds with respect to one another (figure 2.6). The axis system created within the magnetic field is represented by A_{xx}, A_{yy}, and A_{zz} in figure 2.5 and S_{xx}, S_{yy}, and S_{zz} in figure 2.6. The angles displayed in the figure define the orientation of the vector with respect to the magnetic field. The NMR pulse program most commonly used to measure the one-bond amide dipolar coupling represents a modification of the ^{15}N-HSQC. In this type of modified HSQC, the magnetic dipole coupling between the ^{1}H and ^{15}N is allowed to evolve as the ^{15}N chemical shift is observed. Measurement of the amide dipole coupling depends, in part, on the distance between the bonded nitrogen and hydrogen atoms. Since this distance is known, the RDC measurement can be used to establish the angle between the N-H bond and the alignment tensor, which then identifies the relative orientations of the amide bonds with respect to each other, as seen in figure 2.6.

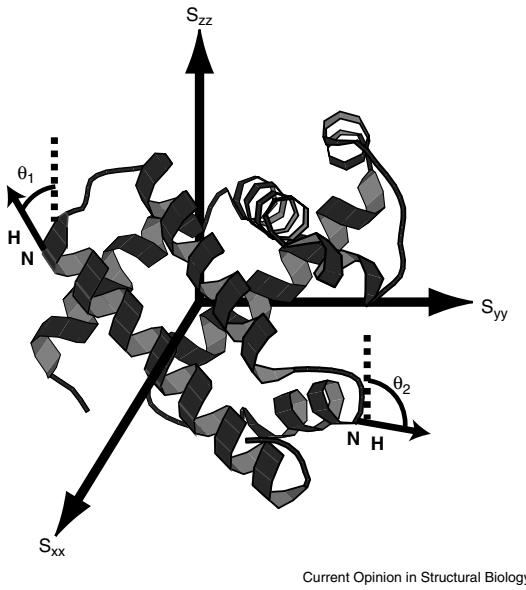


Figure 2.6: Orientation of two dipolar coupling vectors in a protein segment and the axis system in a magnetic field.

Again, the orientation of the molecular axis system in the magnetic field along with the known distance between the two atoms can be used to define the orientation of the bond vector in that axis system. As amino acids are sequentially assigned, the bond vectors are sequentially identified in a manner that allows for structural refinement. Aside from structural refinement, RDC analysis can be used to identify amino acid residues that experience conformational exchange, collect data on binding orientation of ligands in complex with protein, and determine the global fold of proteins despite few NOEs collected (32).

RDCs can be used to both validate and improve existing structure quality. Regarding the latter, RDC measurement can reduce the number of ensemble structures that satisfy all experimental restraints and allow for determination of more accurate structures, especially when few NOEs are available. For RDC-derived structures, the *Q factor* represents structure quality. That is, the agreement between the calculated structure and the experimental data; in this sense the *Q* factor is analogous to the *R* factor used in

x-ray crystallography. Values of approximately 25% represent high-resolution structural refinement.

2.3 EXPERIMENTAL PROCEDURES

2.3.1 Purification of proteins and DNA:

The methylcytosine-binding domain of MBD3 (amino acid residues 1-70) was cloned, expressed and purified as previously described for cMBD2 (18). 17-bp and 27-bp complementary oligonucleotides (Table 1) were purchased (Integrated DNA Technologies), annealed and purified as previously described (18). The sequences were derived from the $p16^{INK4a}$ promoter known to be a native target sequence for MBD2 (5).

2.3.2 NMR spectroscopy:

Purified protein was combined with 10% excess purified dsDNA and buffer exchanged into 10 mM NaPO4, pH 6.5, 1mM dithiothreitol, 10% $^{2}H_2O$ and 0.02% sodium azide and concentrated to 0.2-1.0 mM. NMR spectra from standard experiments for resonance assignments, distance and torsional angle restraints were collected on a Bruker Avance III 700 MHz instrument. Data were processed using NMRPipe (19) and analyzed with CcpNmr (20). Residual dipolar couplings were measured for complexes containing ^{2}H, ^{13}C, ^{15}N labeled protein using standard IPAP experiments and samples aligned by adding ~12mg/mL Pf1 bacteriophage (Asla Biotech, Ltd.). For each aligned sample, a 1D ^{2}H spectrum of $^{2}H_2O$ was collected and the deuterium quadrupole splitting measured. When comparing RDC values between samples, the observed RDC values

were normalized to an effective deuterium quadrupole splitting of 10 Hz.

2.3.3 Structure Calculations:

The structure of the MBD3 MBD was calculated by simulated annealing as implemented in the Xplor-NIH software package (21) and based on NOE derived distance constraints, torsion angle restraints, and residual dipolar couplings as well as a torsion angle database potential of mean force (22) and a quartic van der Waals repulsion term for non- bonded contacts (23). Backbone torsional angle restraints were derived from chemical shifts using the TALOS+ software (24) and hydrogen bond distance and angle restraints were introduced based on backbone torsional angles and characteristic NOE patterns.

2.3.4 Binding affinity:

Binding affinities were determined by surface plasmon resonance analysis on a Biacore T100 system (GE Healthcare) as described previously (18). The binding affinity was determined from steady state analysis of the SPR relative response at varying concentrations of protein. As previously shown, the maximum steady state response (R_{max}) in SPR depends on stoichiometry (n) of binding (25). Prior to fitting, the steady state response at each protein concentration (R_A) was normalized (R_{norm}) to the total DNA immobilized (R_l) and molecular weights of the DNA and protein (MW_L and MW_A, respectively).

$$R_{norm} = \frac{R_A}{R_L \bullet \left(\frac{MW_A}{MW_L} \right)} \qquad (1)$$

Final data analysis, plotting, and curve fitting were performed with pro Fit software (QuantumSoft).

2.4 RESULTS

2.4.1 Solution structure of the MBD3 MBD is nearly identical to that of MBD2.

We determined the solution structure of the MBD from MBD3 (amino acids 1-70) bound to a 17-bp dsDNA with a central hydroxymethylated CpG dinucleotide. The structure was calculated based on 528 NOE derived distance constraints, 120 dihedral angle restraints, and 53 residual dipolar coupling restraints (Table 2.1). The overall protein structure is well defined (Fig. 2.7A-B) with average pairwise root mean square deviations (RMSD) of 0.7±0.1 Å (backbone) and 1.2±0.1 Å (all heavy atoms) for ordered regions (residues 6-23,34-71). As expected, the fold is very similar to that of chicken MBD2 (cMBD2; RMSD=2.0±0.1 Å) for the same ordered regions (Fig. 2.7D) (18).

As with all MBD proteins studied to date, the topology of the MBD comprises a four-stranded β-sheet followed by a single α-helix and a C-terminal loop. The central two strands of the β-sheet (β2 and β3) form a long finger-like projection that can extend down and across the major groove of DNA to make base specific contacts. The most notable difference between the cMBD2 and MBD3 structures is that the loop connecting the long fingerlike projection is not as well ordered in MBD3 (residues 24-33, Fig. 2.7A,C) with

an RMSD of 1.7±0.5 Å (backbone) as compared to cMBD2 with an RMSD of 0.9±0.3 Å. Residues at the base of this loop form critical DNA specific contacts and a hydrogen bond network that stabilizes the interaction with the methylated CpG dinucleotide (Fig. 2.7E).

Predicting backbone order parameters (S^2) based on chemical shifts using the random coil index (RCI) method (24, 26) confirms that this loop is less well ordered in MBD3 (Fig. 2.8). The difference in predicted S^2 between complexes and shows that residues 24-33 become progressively more ordered between the MBD3:hmCpG, MBD3:mCpG and cMBD2:mCpG complexes, respectively. Thus in the absence of a methylation-specific binding mode, the loop connecting the central two β-strands is more flexible.

Filtered intermolecular NOE spectra did not contain NOE crosspeaks, which is consistent with the observed line broadening for residues at the DNA interface and the overall lower affinity of MBD3 for DNA and suggests non-specific protein-DNA interaction with dynamic exchange between binding sites. Based on subsequent analyses that indicate MBD3 preferentially localizes to methylated sites, we collected filtered NOE spectra for MBD3 bound to methylated DNA (mCpG). Likewise, we did not detect any intermolecular NOE crosspeaks with methylated DNA indicating that MBD3 exchanges among different binding modes whether on methylated or hydroxymethylated

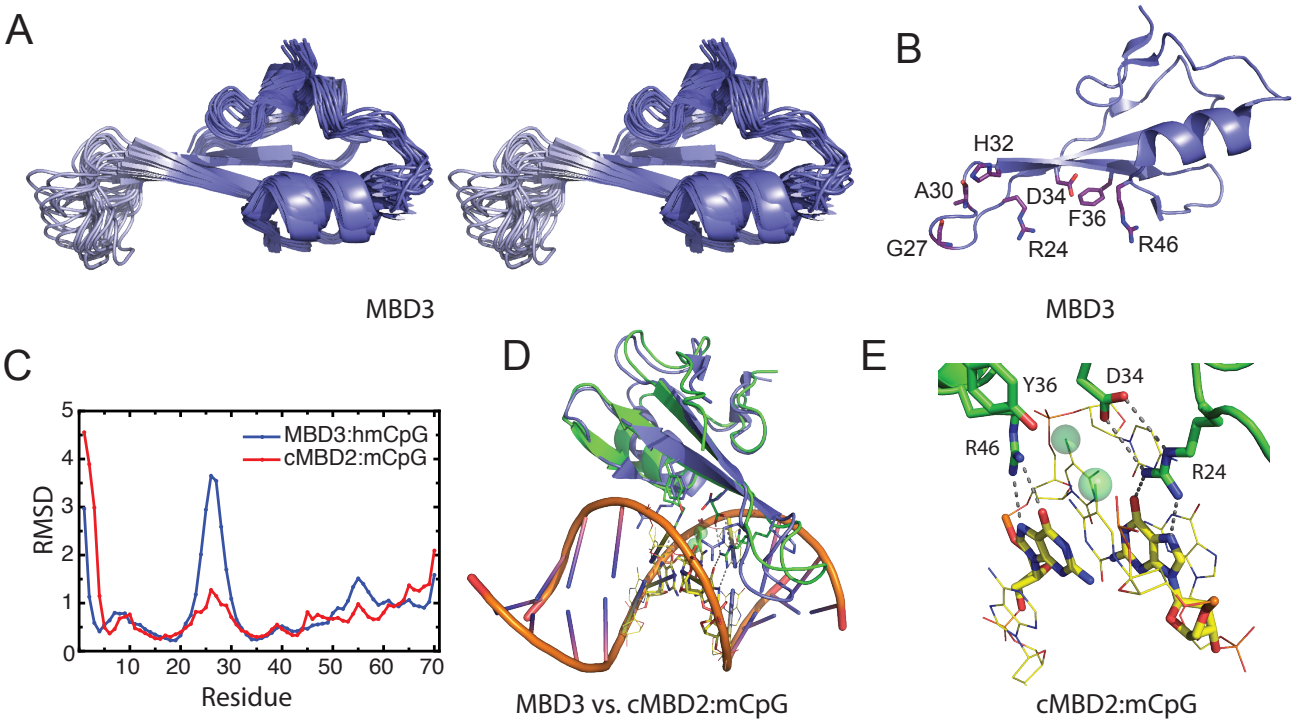


FIGURE 2.7: Solution structure of MBD3 methyl binding domain bound to hydroxymethylated DNA. (A) Stereo ribbon diagram (blue) of the MBD3 solution structure is shown for the ensemble of twenty calculated structures (PDB ID: 2mb7). The loop connecting β2 and β3 (residues 24-33) is highlighted in light blue. (B) Ribbon diagram of the lowest energy solution structure is shown with key contact and chemical shift reporter residues depicted in sticks. (C) Per residue root mean standard deviation (RMSD) for backbone atoms is plotted for the solution structure ensemble of MBD3 (blue) and for the solution structure ensemble of cMBD2 (red) previously reported (PDB ID: 2ky8).(18) (D) The best fit protein alignment of the solution structures of cMBD2 (green) and MBD3 (blue) MBD is shown bound to the methylated DNA from the cMBD2:dsDNA solution structure (PDB ID: 2ky8). (E) Diagram highlighting the cMBD2 hydrogen-bonding network while bound to methylated DNA and with key residues depicted as sticks. Structure diagrams were generated using the Pymol program (Delano

DNA. Given the absence of informative intermolecular NOEs we did not determine a solution structure of the protein-DNA complex. Importantly, we noted that chemical shifts of MBD3 bound to hmCpG (discussed in detail below) are more similar to MBD3 on unmethylated DNA than methylated DNA, which led us to compare spectra of MBD3 on different DNA molecules to probe both methylation-specific and non-specific DNA association.

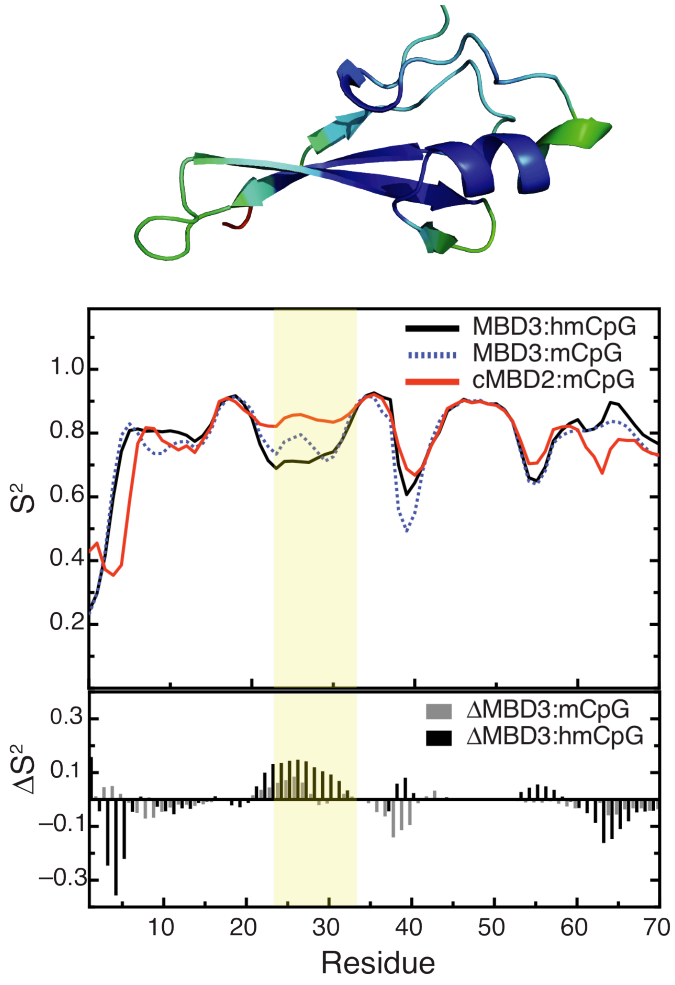


FIGURE 2.8: Methyl-specific binding mode stabilizes a dynamic loop in MBD3. (A) Ribbon diagram of the MBD3 solution structure is shown and colored based on order parameters predicted from chemical shift index (S2 – shading from blue to red reflects low to high). (B) The predicted order parameters (S2) are plotted for the MBD3:hmCpG (black), MBD3:mCpG (blue dotted), and cMBD2:mCpG (red) complexes. (C) Bar plots are shown for the difference in order parameters (ΔS2) between the cMBD2:mCpG complex and MBD3:mCpG (black) and MBD3:hmCpG (gray) complexes. The loop connecting β2 and β3 (residues 24-33) is highlighted in light yellow in plots (B) and (C).

2.4.2 MBD3 spends a significant portion of time on methylated sites.

As first described for MeCP2 (27), and later described for cMBD2, (18) a pair of highly conserved arginine residues in MBD proteins form bi-dentate hydrogen bonds with the symmetrically opposed guanine bases of an mCpG dinucleotide (Fig. 2.7E). The aliphatic portion of each arginine packs against the methyl group of the neighboring

methylcytosine. Given that an unmethylated CpG contains symmetrically opposed guanines, we hypothesized that MBDs should still recognize the CpG dinucleotide, but with lower affinity. Even though MBD3 binds DNA with lower overall affinity and shows less selectivity for methylated DNA, the critical arginine residues are conserved and could provide sequence specific recognition of CpG dinucleotides.

In ^{15}N-HSQC spectra, we noted that several crosspeaks that showed unusual chemical shifts for cMBD2:mCpG did not show the same chemical shifts for MBD3:hmCpG (Fig. 2.9B). In particular, the ^{1}Hε of R24, which forms a sidechain hydrogen bond with D32 and is shifted far downfield (9.5ppm) in cMBD2:mCpG, is only shifted to 7.5ppm in MBD3. Likewise, G27 is shifted upfield in ^{15}N to 102ppm in cMBD2:mCpG, but not to the same degree in MBD3:hmCpG (105ppm), and finally A30 is shifted upfield in ^{1}H to 6.8ppm in cMBD2:mCpG but only 7.6ppm in MBD3:hmCpG.

We hypothesized that these large chemical shift changes reflect a difference between methylation-specific and non-specific binding modes. R24 and D32 are positioned at the N and C terminus of the poorly structured loop in MBD3, while A30 and G27 also fall within this same loop. Methylation-specific binding stabilizes the R24:D32 H-bond and the loop containing G27 and A30. To test this hypothesis, we compared HSQCs for MBD3 bound to a DNA sequence with three CpG dinucleotides in which the central CpG is methylated (mCpG), hydroxymethylated (hmCpG), or unmethylated (CpG(x3)), as well as similar DNA sequences with only one (CpG(x1)) or no CpGs (CpG(x0)) (Table 2.2).

	Protein
NMR distance and dihedral constraints	
Distance constraints	
Total NOE	528
Intra-residue	102
Inter-residue	
Sequential ($\|i-j\|=1$)	155
Medium-range ($\|i-j\|\leq 4$)	111
Long-range ($\|i-j\|>5$)	160
Hydrogen bonds	32
Total dihedral angle restraints	120
ϕ	54
ψ	54
$\chi 1$	12
Total RDCs	
NH	53
Q%	
NH	7.0
Structure statistics	
Violations (mean and s.d.)	
Distance constraints (Å)	0.018±0.003
Dihedral angle constraints (º)	0.4±0.1
Max. dihedral angle violation (º)	4.8
Max. distance constraint violation (Å)	0.48
Deviations from idealized geometry (mean and s.d.)	
Bond lengths (Å)	0.04±0.01
Bond angles (º)	0.69±0.002
Impropers (º)	2.01±0.006
Average pairwise r.m.s. deviation* (Å)	
Heavy	1.2±0.1
Backbone	0.7±0.1
Ramachandran plot summary for ordered residues	
Most favored regions	92.4%
Additionally allowed regions	7.4%
Generously allowed regions	0.2%
Disallowed regions	0.0%

*Pairwise r.m.s. deviation and s.d. from the mean was calculated among 20 lowest energy (out of 50) refined structures for ordered residues (6-23, 34-71).

Table 2.1: NMR and refinement statistics. The number and type of structural constraints as well as the final refinement statistics are presented for the solution structure of MBD3 bound to hydroxymethylated DNA.

We found that the chemical shifts of these reporter resonances fall on a line between extrema represented by cMBD2:mCpG and MBD3:CpG(x0). Importantly, the peak position for each of the reporter resonances falls at the same fractional distance between these extrema (Fig. 2.9B), shifting toward the position in the cMBD2:mCpG complex as the number of unmethylated CpG sites increases and with the addition of a methylated CpG. These observations strongly indicate chemical shift averaging between two binding modes (28) reflective of fast exchange between methylation-specific and non-sequence specific interaction with DNA. Furthermore, the chemical shift changes are consistent with preferential localization at the CpG and mCpG sites.

Using cMBD2 as representative of the methylation-specific binding mode could introduce structural and primary sequence differences that affect observed chemical shifts. Therefore, we sought to generate an MBD3 MBD that binds with high selectivity for mCpG, which would allow us to evaluate chemical shift changes for more backbone resonances with fewer confounding sequence variations. Previous studies have established that the lack of mCpG specificity for MBD3 reflects two amino acids (H30 and F34) that differ from other MBD proteins (K32 and Y36 in cMBD2) (4, 5). We introduced the H30K, F34Y double mutation into MBD3 ($MBD3^{KY}$) and, as expected, this mutant bound with higher affinity and selectivity for mCpG comparable to cMBD2 (Fig. 2.13, Table 2.3). Each of the reporter resonances now show chemical shifts that are very similar to those of the cMBD2:mCpG complex confirming that the unique chemical shifts do reflect a methylation-specific binding mode.

Name	Length (bp)	Sequence
mCpG	17	GAGGCGCT(mC)GGCGGCAG
hmCpG	17	GAGGCGCT(hmC)GGCGGCAG
CpG(x3)	17	GAGGCGCTCGGCGGCAG
CpG(x1)	17	GAGGCCCTCGGGGGCAG
CpG(x0)	17	GAGGCCCTCTGGGGCAG
$mCpG^{27}$	27	GAGCTAGAGCGCT(mC)GGCGGCGCCAGGC
$mCpG^{10}$	10	GGAT(mC)GGCTC

Table 2.2: DNA sequences. The length and nucleotide sequences are given for the different dsDNA molecules used for MBD2 and MBD3 binding studies.

To further explore the chemical shift changes associated with the different binding modes, we assigned the backbone resonances (^{15}N, ^{1}H) for MBD3 bound to mCpG or CpG(x3) and $MBD3^{KY}$ bound to mCpG. In Figure 3A, the chemical shift distances are plotted for backbone resonances between the different complexes. The largest chemical shift changes are seen for the poorly structured loop (residues 24-33) when comparing the $MBD3^{KY}$:mCpG and MBD3:mCpG complexes with the MBD3:hmCpG complex. In contrast, the ^{15}N-HSQC spectra for the MBD3:hmCpG and MBD3:CpG(x3) complexes show nearly identical chemical shifts and spectra (Fig. 2.11A).

These observations strongly support a binding model in which MBD3 exhibits fast exchange between methylation-specific and non-specific binding. Thus the observed chemical shift (σ_{obs}) reflects a weighted average of the methylation-specific ($\sigma_{MBD3(KY):mCpG}$) and non-specific ($\sigma_{MBD3:CpG(x3)}$) binding modes, as given by:

$$\sigma_{obs} = \rho_m \bullet \sigma_{MBD3^{KY}:mCPG} + (1-\rho_m) \bullet \sigma_{MBD3:CPG(x3)} \quad (2)$$

where ρ_m is the fraction in the methylation-specific binding mode. Hence, these chemical shifts are a direct measure of the average time spent on the mCpG site. Using eleven backbone ^{1}H, ^{15}N resonances that show a clear linear relationship between chemical shift

and DNA bound, we find that MBD3 spends approximately 43% (±5%) of the time on the mCpG site (Fig. 2.9B).

These data further indicate that, despite a lack of a strong global binding affinity preference for mCpG DNA, MBD3 still spends a significant proportion of time on methylated sites. To confirm that this finding was not the result of very weak binding with chemical shift averaging between DNA bound and free MBD3, we compared HSQCs for 600μM and 300μM samples of protein on DNA. The peaks for each reporter residue show nearly identical chemical shifts at both concentrations (Fig. 2.11B), indicating that the observed chemical shift changes of the reporter residues were not the result of exchange between bound and free states but instead represent averaging between different bound states. The observed differences in chemical shift represent changes in the binding distribution on DNA, not changes in the distribution between bound and free states.

We developed a statistical-mechanical model to describe the distribution of MBD3 on methylated DNA (Fig. 2.9C) in which the partition function comprises a sum of Boltzmann factors ($e^{-E(i)/kT}$) for methylation-specific and non-specific binding modes, ΔE is the difference in energy between binding modes and N is the number of non-specific sites. The additional methylation- specific interactions formed by MBD2 lead to a larger ΔE thereby increasing ρ_m and the relative selectivity for mCpG. Without these interactions, MBD3 shows a smaller ΔE, ρ_m is reduced but still non-zero, and MBD3 distributes unevenly between mCpGs and unmethylated sites. Figure 2.9C depicts a simplified structural model of this distribution in which the MBD3 structure was docked onto the 17-bp DNA fragment at the centrally methylated mCpG (red) as well as four

"non-specific" binding sites (blue). The statistical model further indicates that localizing to a mCpG depends on the number of non-specific sites available (N), and therefore the length of DNA. We tested this latter prediction by comparing chemical shifts for the reporter residues of MBD3 on 10, 17, and 27-bp DNA with a single mCpG (Fig. 2.9D). Indeed, we found that increasing DNA length results in chemical shift changes for G27 and A30 toward the unmethylated binding mode.

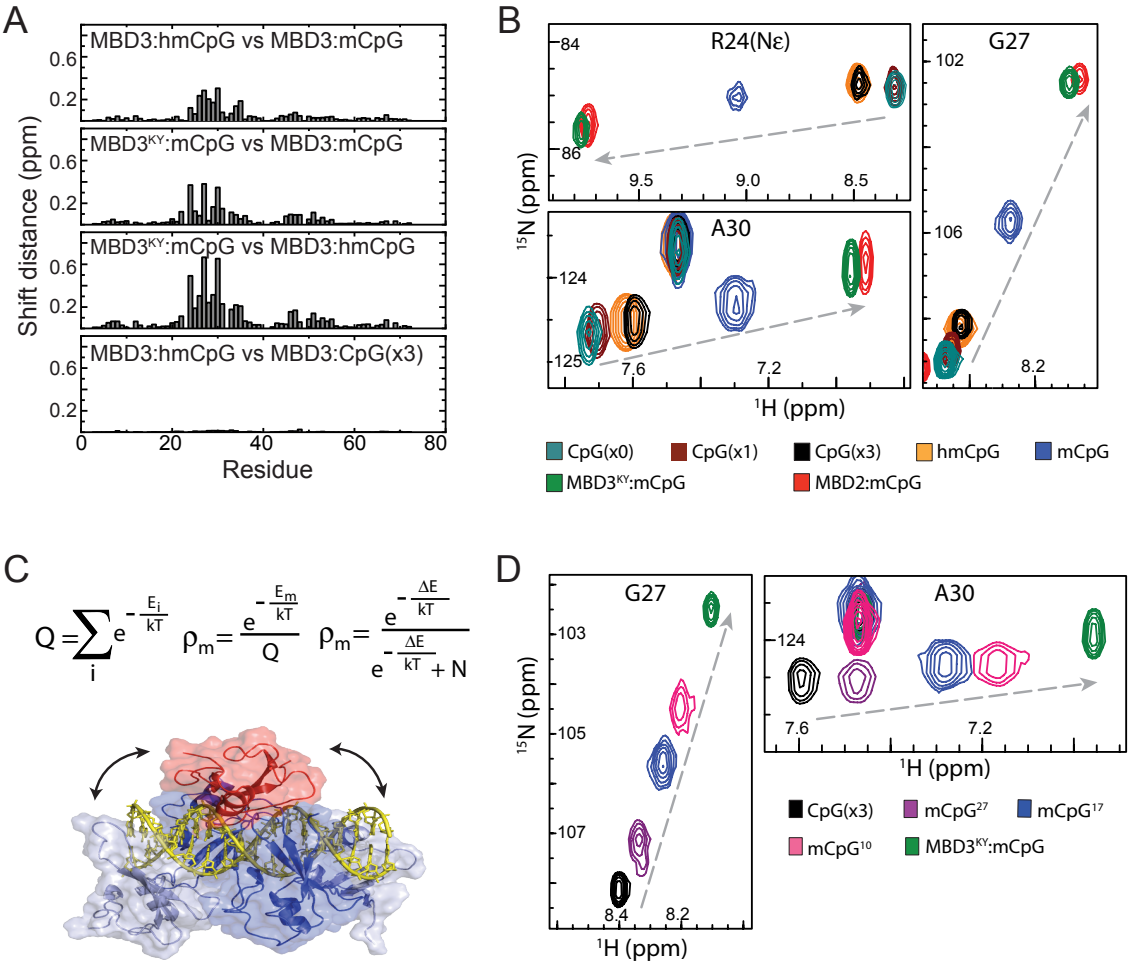


FIGURE 2.9: Preferential localization of MBD3 to mCpG sites. (A) Bar plots show the chemical shift distances between MBD3:dsDNA complexes. (B) An overlay of 15N-HSQC spectra are shown for key reporter residues of MBD3 bound to CpG(0x), CpG(1x), CpG(3x), mCpG and hmCpG as well as MBD3KY and cMBD2 bound to mCpG. (C) The derivation of a simple statistical mechanical model for the distribution of MBD3 on mCpG (top) is shown with a mixed rendering diagram (bottom) depicting MBD3 docked onto a methylated site (red) as well as four non-methylated sites (blue) of the mCpG DNA. Arrows indicate rapid exchange between these binding modes. (D) Overlay of 15N-HSQC spectra are shown for key reporter residues of MBD3 and MBD3KY while bound to DNA of varying lengths.

2.4.3 Residual dipolar couplings confirm MBD3 localizes to methylated sites without significant conformational change:

As an alternative method to assess the ensemble of binding modes, we measured residual dipolar couplings ($^1D_{NH}$) for MBD3 bound to methylated and unmethylated DNA as well as for MBD3^{KY} bound to methylated DNA. As can be seen in Figure 2.10, the observed $^1D_{NH}$ are similar but not identical between the different complexes. When plotting $^1D_{NH}$ from the unmethylated complex against those from wild type MBD3 or MBD3KY methylated complexes, the data fall off of the line of identity (y=x), as highlighted by the red ovals in Figure 2.10A-B. Since the observed residual dipolar couplings reflect a weighted average of the different binding modes, (29) the $^1D_{NH}$ for each residue (n) of MBD3 bound to mCpG is a linear combination of $^1D_{NH}$ for MBD3 bound to CpG(x3) (non-specific binding mode) and $^1D_{NH}$ for MBD3^{KY} bound to mCpG (methylation-specific binding mode):

$$^1D^{pred}_{NH,MBD3:mCpG}(n) = \rho_m \bullet {}^1D^{obs}_{NH,MBD3^{KY}:mCpG}(n) + (1-\rho_m) \bullet {}^1D^{obs}_{NH,MBD3:CpG(x3)}(n) \quad (3)$$

In Eq. (3), ρ_m is the fraction in the mCpG specific binding mode and $^1D_{NH}(n)$ are the residual dipolar couplings for each residue in the MBD3:mCpG, MBD3:CpG, and MBD3KY:mCpG complexes. Fitting the observed $^1D_{NH}$ to Eq. (3) as a function of ρ_m provides the fraction of MBD3 in the methylation-specific binding mode. As shown in Fig 2.10C, the sum of squared residuals (SSR) between predicted and observed $^1D_{NH}$ is minimized when ρ_m is 0.37. Plotting observed $^1D_{NH,MBD3:mCpG}$ versus predicted at ρ_m = 0.37 shows good agreement with tight clustering around y=x (Fig. 2.10D). Therefore, using RDCs as an independent measure of methylation selectivity, we find that MBD3

spends approximately 37% of the time on the methylated binding site, which agrees within experimental error with the results from chemical shift analysis.

The residual dipolar couplings for each complex were fit to the solution structure of MBD3 bound to hmCpG using singular value decomposition as implemented by the PALES software (30). Despite the differences in RDCs between complexes, each data set fit quite well to the MBD3 structure with *Q* factors of 19.9%, 23.1%, and 25.9% for the MBD3:mCpG, MBD3:CpG(x3), and MBD3^{KY}:mCpG complexes, respectively (Fig. 2.10E). Therefore, the change between methylated and non-specific binding modes does not involve backbone structural changes. The observed chemical shift changes likely reflect a stabilization of the dynamic loop between β2 and β3 but without significant structural rearrangements.

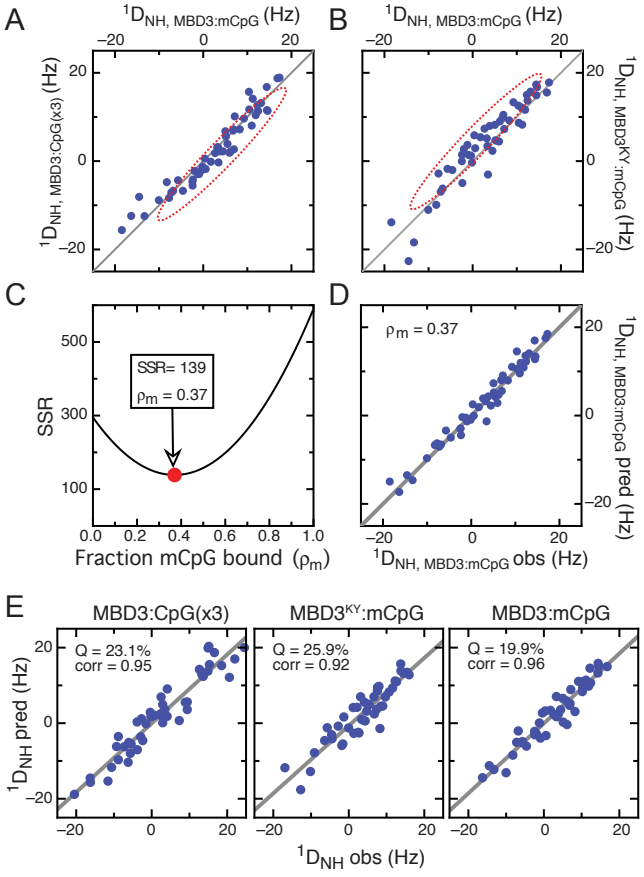


FIGURE 2.10: MBD3 localizes to methylated DNA sites without significant conformational change. A comparison of measured $^{1}D^{NH}$ RDCs normalized to $^{2}H2O$ quadrupole splitting of 10 Hz are plotted for (A) MBD3:mCpG vs MBD3:CpG(x3) and (B) MBD3:mCpG vs MBD3KY:mCpG complexes. Red ovals highlight those values that fall of the line of identity (gray line). (C) The sum of squared residuals (SSR) is plotted as a function of ρm (Eq. (3)). The SSR is minimized (red circle and arrow) at 37% mCpG bound (ρm = 0.37). (D) Plotting $1D^{NH}$ RDCs for MBD3:mCpG observed vs. predicted with ρm = 0.37 ((3)) shows good agreement with tight clustering around y = x. (E) The measured $1D^{NH}$ RDCs for each complex (MBD3:CpG(x3), MBD3KY:mCpG, and MBD3:mCpG, left to right plots, respectively) were fit to the solution structure of MBD3 by SVD and the observed vs. predicted values plotted. The *Q* factors and correlation coefficients show good agreement with the solution structure indicating that the backbone structure of MBD3 does not change between complexes.

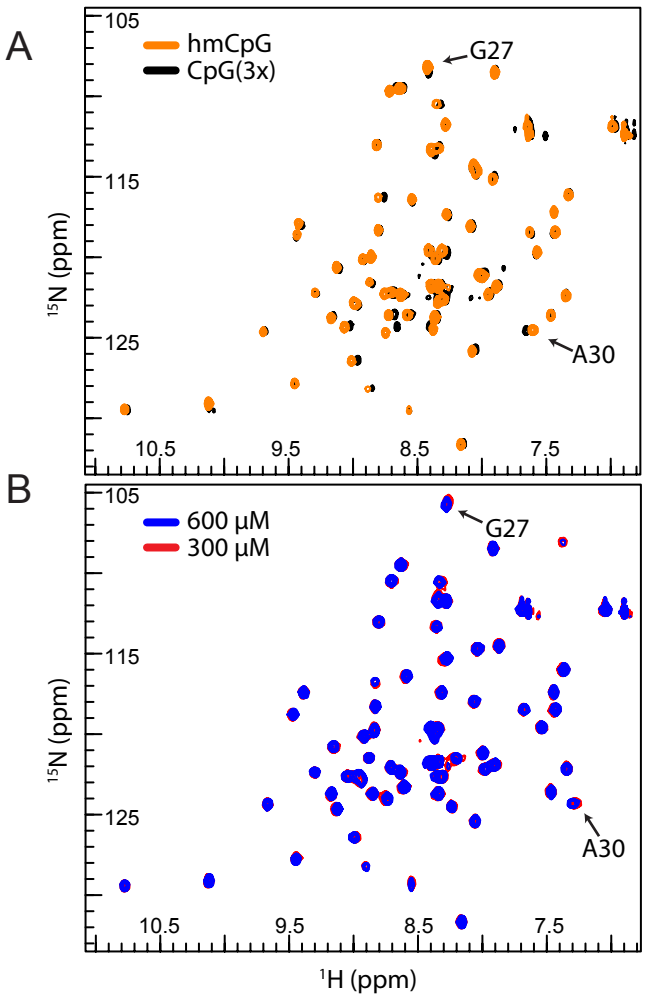
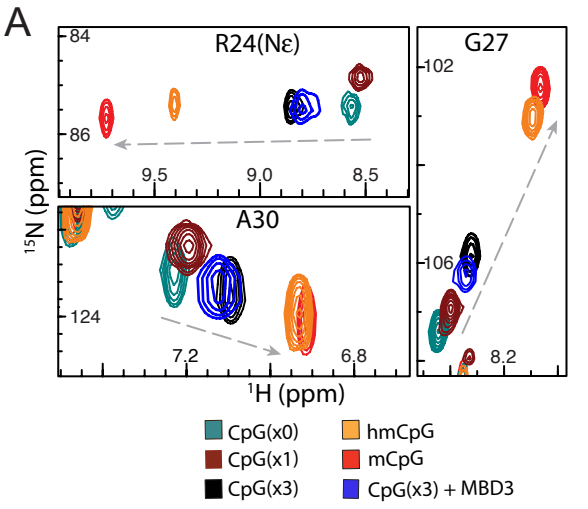


FIGURE 2.11: Chemical shifts do not depend on concentration or the presence of hydroxymethlyation. Overlay of ^{15}N-HSQC spectra are shown comparing a) MBD3:hmCpG (orange) with MBD3:CpG(3x) and b) MBD3:mCpG at 600 µM (blue) and 300 µM (red). Resonances for key reporter residues A30 and G27 are labeled.

2.4.4 MBD2 and MBD3 distribution is influenced by DNA methylation status and CpG density:

In contrast to MBD3, MBD2 appears to spend most of its time on the methylated site (Fig. 2.12A). The chemical shifts for reporter residues in MBD2 represent the extrema for the complexes studied. As shown in Figure 6B, these peaks do not change with increasing DNA length, indicating that a large ΔE dominates the fraction bound to the methylated site and $\rho_m \approx 1$ (Fig. 2.9C). The higher affinity and methylation selectivity of MBD2 results in exclusive occupancy of mCpG.

Comparing spectra of MBD2 and MBD3 when bound to the unmethylated oligonucleotides CpG(x0), CpG(x1) and CpG(x3) reveals that the reporter resonances shift towards the methylation-specific state with increasing numbers of CpGs (Fig. 2.9B and 2.13A). This finding indicates that both MBD2 and MBD3 localize to sites of increased CpG density. MBD3 shows relatively small chemical shift changes when bound to CpG(x3) as compared to CpG(x0), consistent with localizing to a CpG dinucleotide approximately 9% (±3%) of the time. MBD2 shows a pronounced difference between CpG(x0) and CpG(x3) (Fig. 2.12A) showing that MBD2 preferentially localizes to CpG dinucleotides approximately 33% (±11%) of the time. The reporter resonances also indicate that MBD2, but not MBD3, tends to localize to hmCpG and exhibits higher affinity for these sites (Fig. 2.12A, Table 2.3). To test whether MBD3 could influence how MBD2 distributes on unmethylated CpGs, we added equimolar MBD3 to ^{15}N-MBD2 bound to CpG(x3) and found that the reporter peaks shifted towards those of the non-specific binding mode (Fig 2.12A). This finding shows that despite relatively weak overall binding affinity, MBD3 can modulate the distribution of MBD2 on CpG sites, shifting MBD2 towards the non- specific binding mode.



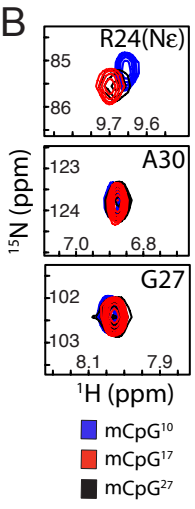


FIGURE 2.12: MBD2 distribution is influenced by DNA methylation status and CpG density. (A) An overlay of ^{15}N-HSQC spectra for key reporter residues of MBD2 bound to CpG(0x), CpG(1x), CpG(3x) +/- MBD3, mCpG and hmCpG, shows that cMBD2 preferentially localizes to DNA with mCpG, hmCpG and multiple CpG sites, and that localization is modified by the presence of equimolar MBD3. (B) An overlay of ^{15}N-HSQC spectra cMBD2 bound to mCpG DNA of varying lengths confirms that cMBD2 strongly prefers mCpG sites.

2.4.5 Global binding affinity does not reflect localization preferences on DNA:

Binding affinities for each of the methylated and unmethylated sequences investigated were determined by steady-state analysis of surface plasmon resonance data, as described previously (18). To allow for direct comparison of binding stoichiometry, we normalized the relative steady-state response to the amount of DNA coupled to the sensor chip such that the maximum steady-state response reflects the number of binding sites on the DNA (25). As can be seen in Fig. 2.13 and Table 2.3, both cMBD2 and $MBD3^{KY}$ bind mCpG DNA with high affinity and stoichiometry of approximately one (K_D = 105 ± 7 nM and113 ± 11nM and R_{max} = 0.83 ± 0.01 and 0.98 ± 0.02, respectively). Both bind unmethylated DNA with much lower affinity and high stoichiometry (K_D = 74 ± 4 μM and 17 ± 2 μM and R_{max} = 4.8 ± 0.1 and 2.1 ± 0.1, respectively) indicative of a high degree of methylation selectivity.

While NMR analyses indicate MBD3 binds DNA with sufficient affinity to be fully bound at 300μM concentration (Fig. 2.11B), solubility limits of the isolated protein preclude accurate determination of binding constants by surface plasmon resonance. The qualitative results of these studies, however, are very similar to those reported previously by Hashimoto *et al. (15)* who determined DNA binding affinity for full-length MBD3 by fluorescence anisotropy. In those previous studies, MBD3 bound with similar low affinity to unmethylated and hydroxymethylated DNA and with a small but weak preference (approximately five-fold) for methylated DNA. The current SPR data indicates a small preference for mCpG as well (Fig. 2.13).

Likewise, cMBD2 shows similar low affinity and high stoichiometry when binding to hydroxymethylated and unmethylated DNA with 0-3 CpG dinucleotides (Fig. 2.13 and

Table 2.3). Therefore, global binding analysis reveals a marked preference for mCpG by cMBD2 and MBD3^{KY} but only weak and non-specific binding for all other complexes. These findings agree with several previous DNA binding analyses of MBD proteins (4, 5, 15, 18).

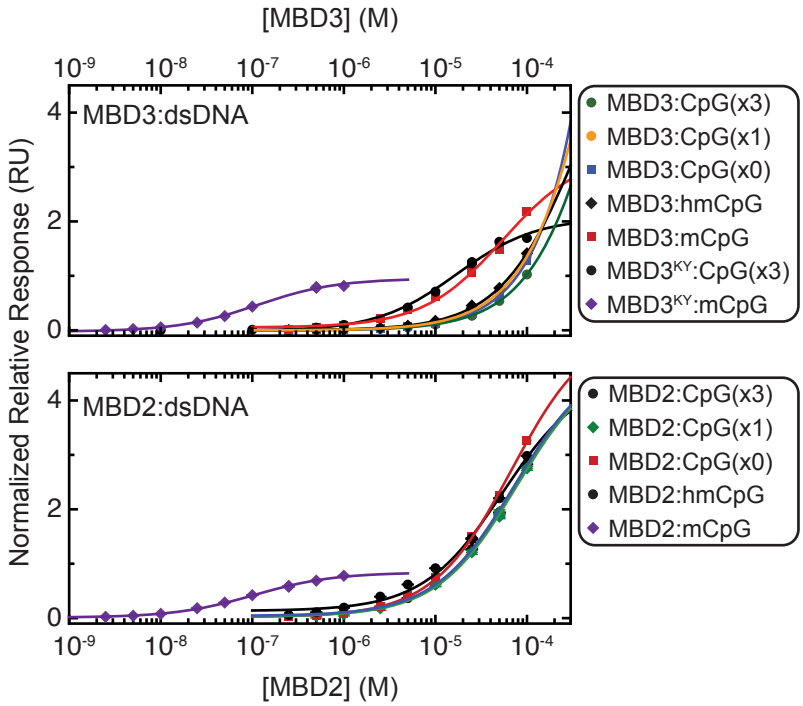


FIGURE 2.13: MBD2 and MBD3 localization does not translate into a global binding affinity preference. Steady state surface plasmon resonance measurements are shown for MBD3 (top panel) and cMBD2 (bottom panel) binding to immobilized double stranded oligonucleotides of varying CpG content and methylation status. The steady-state response was normalized to the amount of DNA immobilized (Eq. (1)) such that the maximum response reflects the stoichiometry of binding (n).

Protein	DNA	K_D (μM) ± SE	R_{max}*	Chi^2 ($x10^{-3}$)
MBD3	mCpG	54 ± 7	3.2	25
MBD3	CpG(x3)	n.d.**	-	-
MBD3	CpG(x1)	n.d.	-	-
MBD3	CpG(x0)	n.d.	-	-
MBD3	hmCpG	n.d.	-	-
MBD3^{KY}	mCpG)	0.13 ± 0.01	1.0	2.6
MBD3^{KY}	CpG(x3)	17 ± 2	2.1	1.8
MBD2	mCpG	0.11 ± 0.01	0.8	1.0
MBD2	CpG(x3)	74 ± 4	4.8	3.8
MBD2	CpG(x1)	78 ± 7	4.8	7.2
MBD2	CpG(x0)	68 ± 3	5.4	5.9
MBD2	hmCpG	54 ± 8	4.3	51

*Normalized such that R_{max} reflects stoichiometry (n)
**Weak binding and solubility limits preclude accurate determination of binding affinity.

Table 2.3: Binding affinity. The dissociation constant (K_D), R_{max}, and Chi^2 are given for different protein and DNA complexes as determined by steady state analysis of surface plasmon resonance studies.

2.5 DISCUSSION

Although it has been established that MBD3 binds DNA with lower affinity and much less specificity for mCpG dinucleotides than MBD2, the functional role of MBD3 has not been well defined. Based on NMR structural, chemical shift, and residual dipolar coupling analyses, we have demonstrated that MBD3 binding to methylated DNA can be described by an ensemble of methylation-specific and non-specific binding modes and that MBD3 preferentially localizes, albeit weakly, to methylated sites. In contrast to MBD2, MBD3 does not distinguish between hydroxymetylated and unmethylated DNA but does show a slight preference for DNA with multiple CpG dinucleotides.

A similar chemical shift analysis of cMBD2 showed that cMBD2 localizes almost exclusively to the methylated sites when present. Even on unmethylated DNA, cMBD2 preferentially localizes to CpG dinucleotides, especially as the number of available CpGs increases. Surprisingly, cMBD2 more avidly localized to a hydroxymethylated than an unmethylated site. Taken together, these findings lead to a model in which both MBD2 and MBD3 preferentially localize to DNA with multiple CpGs while MBD2 more exclusively localizes to mCpGs.

Our studies lead to several novel observations and hypotheses. First, the differences in DNA binding between MBD3 and MBD2 reflect a change in the degree of selectivity for mCpG, not absolute differences in binding specificity. As we hypothesized, both MBD2 and MBD3 show a weak preference for CpG dinucleotides even the absence of methylation; however, MBD2 shows a much greater selectivity for mCpG than MBD3. These preferences are not necessarily apparent by global binding analyses but instead are

reflected in the partitioning between methylation-specific and non-specific binding modes on DNA, observed via NMR analysis.

The observed changes in the NMR spectra arise from changes in the distribution of bound states, not from changes in the distribution between bound and free states. In addition, the preference, though weak, for multiple CpG sites within the relatively small 17-base pair oligonucleotide used in these studies correlates with the whole genome analyses that show MBD2 and MBD3 localize to CGIs. Given that a CGI contains 100-1000s of CpG dinucleotides, even a relatively small preference would lead to fairly strong localization at such sites. Furthermore, these findings correlate with the observations that both MBD2 and MBD3 are found at unmethylated CGIs, whereas MBD2 binds with much greater affinity and likely excludes MBD3 from methylated CGIs. Finally, these studies clearly indicate that MBD3 does not exhibit a binding preference for or a structural recognition of hmCpG DNA. From the standpoint of MBD3, hydroxymethylation is functionally equivalent to demethylation.

Genomes containing both MBD2 and MBD3 proteins emerge at the same time as the vertebrate methylation pattern, which includes largely unmethylated CGIs. This concurrence along with the preceding characterization of DNA binding by MBD3 leads us to speculate that MBD3 plays an important role in regulating genes with unmethylated CGIs. One possibility these studies raise is that MBD3 helps counterbalance the tendency of MBD2 to preferentially localize to CpG dinucleotides, and consequently to aberrantly silence the associated genes, by competing with MBD2 at unmethylated CGIs.

We found that MBD3 does modify the distribution of MBD2 on DNA such that MBD2 spends less time on CpG sites. Thus MBD3 could help prevent gene silencing by

MBD2 at unmethylated gene promoters and enhancers depending on the relative concentration of the two proteins and cellular context. Indeed, recent studies have shown that knockdown of MBD3 can lead to decreased gene expression (16, 17).

Mammalian cells express multiple MBD proteins as well as different isoforms of individual MBDs. Different splice variants as well as distinct genes encode for MBD2 and MBD3 proteins, some of which lack the DNA binding domain itself. Here, we have studied the solution structure and DNA binding of MBD3 by NMR. Chemical shift analyses indicate that MBD3 recognizes and preferentially localizes to both mCpG and CpG sites but not to the same extent as MBD2. Single amino acid differences dictate the degree to which these proteins localize on mCpG sites. Importantly, these binding characteristics do not necessarily lead to changes in global binding affinities but rather correlate with localization of MBD2 and MBD3 to CGIs in whole cells. Hence, a number of different NuRD complexes can be formed that show varying degrees of DNA methylation selectivity and provide distinct functional roles. For MBD3, these functional differences appear to reflect subtle distinctions in the behavior of the MBD when bound to methylated and unmethylated DNA. Indeed, recent studies have shown that MBD3-NuRD may serve multiple roles in modifying expression for both active and silent genes through localization at promoters, gene bodies, and enhancers of active genes (31). Therefore, these data establish a structural basis for the relative distribution of MBD2 and MBD3 on genomic DNA and help explain their observed occupancy at CpG-rich promoters.

2.6 REFERENCES

1. Hendrich, B., and Bird, A. (1998) Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol.Cell.Biol*. **18,** 6538-6547
2. Hendrich, B., and Tweedie, S. (2003) The methyl-CpG binding domain and the evolving role of DNA methylation in animals. *Trends Genet*. **19,** 269-277
3. Matsumoto, M., and Toraya, T. (2008) cDNA cloning, expression, and characterization of methyl- CpG-binding domain type 2/3 proteins from starfish and sea urchin. *Gene*. **420,** 125-134
4. Saito, M., and Ishikawa, F. (2002) The mCpG-binding domain of human MBD3 does not bind to mCpG but interacts with NuRD/Mi2 components HDAC1 and MTA2. *J.Biol.Chem*. **277,** 35434-35439
5. Fraga, M.F., Ballestar, E., Montoya, G., Taysavang, P., Wade, P.A., and Esteller, M. (2003) The affinity of different MBD proteins for a specific methylated locus depends on their intrinsic binding properties. *Nucleic Acids Res*. **31,** 1765-1774
6. Zhang, Y., Ng, H.H., Erdjument-Bromage, H., Tempst, P., Bird, A., and Reinberg, D. (1999) Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. *Genes Dev*. **13,** 1924-1935
7. Le Guezennec, X., Vermeulen, M., Brinkman, A.B., Hoeijmakers, W.A., Cohen, A., Lasonder, E., and Stunnenberg, H.G. (2006) MBD2/NuRD and MBD3/NuRD, two distinct complexes with different biochemical and functional properties. *Mol.Cell.Biol*. **26,** 843-851
8. Gnanapragasam, M.N., Scarsdale, J.N., Amaya, M.L., Webb, H.D., Desai, M.A.,

Walavalkar, N.M., Wang, S.Z., Zu Zhu, S., Ginder, G.D., and Williams, D.C.,Jr (2011) p66{alpha}-MBD2 coiled-coil interaction and recruitment of Mi-2 are critical for globin gene silencing by the MBD2-NuRD complex. *Proc.Natl.Acad.Sci.U.S.A.* **108,** 7487-7492

9. Denslow, S.A., and Wade, P.A. (2007) The human Mi-2/NuRD complex and gene regulation. *Oncogene*. **26,** 5433-5438
10. Kransdorf, E.P., Wang, S.Z., Zhu, S.Z., Langston, T.B., Rupon, J.W., and Ginder, G.D. (2006) MBD2 is a critical component of a methyl cytosine-binding protein complex isolated from primary erythroid cells. *Blood*. **108,** 2836-2845
11. Singal, R., Wang, S.Z., Sargent, T., Zhu, S.Z., and Ginder, G.D. (2002) Methylation of promoter proximal-transcribed sequences of an embryonic globin gene inhibits transcription in primary erythroid cells and promotes formation of a cell type-specific methyl cytosine binding complex. *J.Biol.Chem*. **277,** 1897-1905
12. Feng, Q., and Zhang, Y. (2001) The MeCP1 complex represses transcription through preferential binding, remodeling, and deacetylating methylated nucleosomes. *Genes & Development*. **15,** 827-832
13. Ng, H.H., Zhang, Y., Hendrich, B., Johnson, C.A., Turner, B.M., Erdjument-Bromage, H., Tempst, P., Reinberg, D., and Bird, A. (1999) MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. *Nat.Genet*. **23,** 58-61
14. Yildirim, O., Li, R., Hung, J.H., Chen, P.B., Dong, X., Ee, L.S., Weng, Z., Rando, O.J., and Fazzio, T.G. (2011) Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. *Cell*. **147,** 1498-

1510

15. Hashimoto, H., Liu, Y., Upadhyay, A.K., Chang, Y., Howerton, S.B., Vertino, P.M., Zhang, X., and Cheng, X. (2012) Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Res*. **40,** 4841-4849
16. Baubec, T., Ivanek, R., Lienert, F., and Schubeler, D. (2013) Methylation-dependent and -independent genomic targeting principles of the MBD protein family. *Cell*. **153,** 480-492
17. Gunther, K., Rust, M., Leers, J., Boettger, T., Scharfe, M., Jarek, M., Bartkuhn, M., and Renkawitz, R. (2013) Differential roles for MBD2 and MBD3 at methylated CpG islands, active promoters and binding to exon sequences. *Nucleic Acids Res*. **41,** 3010-3021
18. Scarsdale, J.N., Webb, H.D., Ginder, G.D., and Williams, D.C.,Jr (2011) Solution structure and dynamic analysis of chicken MBD2 methyl binding domain bound to a target-methylated DNA sequence. *Nucleic Acids Res*. **39,** 6741-6752
19. Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J*. *Biomol*. *NMR*. **6,** 277-293
20. Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J., and Laue, E.D. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins*. **59,** 687-696.

21. Schwieters, C.D., Kuszewski, J.J., Tjandra, N., and Clore, G.M. (2003) The Xplor-NIH NMR molecular structure determination package. *J.Magn.Reson*. **160,** 65-73

22. Clore, G.M., and Kuszewski, J. (2002) Chi(1) rotamer populations and angles of mobile surface side chains are accurately predicted by a torsion angle database potential of mean force. *J.Am.Chem.Soc*. **124,** 2866-2867

23. Clore, G.M., and Gronenborn, A.M. (1998) New methods of structure refinement for macromolecular structure determination by NMR. *Proc.Natl.Acad.Sci.U.S.A*. **95,** 5891-5898

24. Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J.Biomol.NMR*. **44,** 213-223

25. Mistrik, P., Moreau, F., and Allen, J.M. (2004) BiaCore analysis of leptin-leptin receptor interaction: evidence for 1:1 stoichiometry. *Anal.Biochem*. **327,** 271-277

26. Berjanskii, M.V., and Wishart, D.S. (2005) A Simple Method To Predict Protein Flexibility Using Secondary Chemical Shifts. *J.Am.Chem.Soc*. **127,** 14970-14971

27. Ho, K.L., McNae, I.W., Schmiedeberg, L., Klose, R.J., Bird, A.P., and Walkinshaw, M.D. (2008) MeCP2 binding to DNA depends upon hydration at methyl-CpG. *Mol.Cell*. **29,** 525-531

28. Tzeng, S.R., and Kalodimos, C.G. (2012) Protein activity regulation by conformational entropy. *Nature*. **488,** 236-240

29. Williams, D.C.,Jr, Cai, M., Suh, J.Y., Peterkofsky, A., and Clore, G.M. (2005) Solution NMR structure of the 48-kDa IIAMannose-HPr complex of the Escherichia coli mannose phosphotransferase system. *J.Biol.Chem.* **280,** 20775-20784

30. Zweckstetter, M., and Bax, A. (2000) Prediction of Sterically Induced Alignment in a Dilute Liquid Crystalline Phase: Aid to Protein Structure Determination by NMR. *J.Am.Chem.Soc.* **122,** 3791-3792

31. Shimbo, T., Du, Y., et al. (2013) MBD3 Localizes at Promoters, Gene Bodies and Enhancers of Active Genes. *PLoS Genetics*. **9**, e1004028.

32. Tolman, J.R. (2001) Dipolar Couplings as a Probe of Molecular Dynamics and Structure in Solution. *Curr Opin Struct Biol.* **11**, 532-9.

Chapter 3- Sponge MBD2 targets methylated DNA and recruits NuRD components

3.1 ABSTRACT

Invertebrates and vertebrates exhibit substantial variation in DNA methylation patterns. Methyl-cytosine binding domain (MBD) family members also vary in both types of organisms. Whereas invertebrates are known to express only MBD2/3 homologs, vertebrates harbor several MBD family members. Expansion of the MBD family appears to have accompanied the invertebrate-vertebrate transition, though MBD2 preference for methylated DNA has largely been retained throughout the animal lineage. Sponges represent the base of the animal kingdom, and recent genomic analyses of sponges *Amphimedon queenslandica* and *Ephydatia muelleri* indicate the presence of individual genes for MBD2/NuRD complex members; no other MBD family members have been identified in sponge genomes. We questioned whether or not sponge MBD2 targets the NuRD complex to methylated DNA, like other metazoan orthologs, and assessed the ability of MBD2 to select for methylated DNA and recruit NuRD components in the freshwater sponge *Ephydatia muelleri*. We observed that *Ephydatia* MBD2 binds methylated DNA with approximately 150-fold selectivity over unmethylated, compared to a 430-fold preference for methylated DNA exhibited by human MBD2. The *Ephydatia* MBD2 and p66α coiled-coil regions display low identity with human and little pre-formed helical content, but nevertheless, the sponge proteins possess nearly all amino acid residues necessary for high affinity binding observed by the human MBD2-p66α coiled-coils. Despite this similarity, the sponge coiled-coils bind weakly, though

interactions between the sponge and human coiled-coils occur much more avidly. Further, the *Ephydatia* MBD2 p55BR recruits human MTA2, but not human NuRD components HDAC 1/2 and Rbap 46/48, unlike the human p55BR, which can recruit all three NuRD core components. Finally, RNAi-induced knockdown of *Ephydatia* MBD2 results in a low-growth phenotype marked by failure to form structures associated with early development in sponges. Taken together, these data suggest that sponges likely likely harbor a methylation-specific NuRD complex similar to that found in humans and thus have the capacity to carry out sophisticated epigenetic processes, including targeted chromatin remodeling, that likely pre-date sponge emergence.

3.2 INTRODUCTION

Patterns of DNA methylation changed during the transition from invertebrate to vertebrate life. While invertebrate methylomes often harbor "mosaic" methylation patterns in gene bodies, vertebrates tend to show a global methylation pattern that often is observed in promoter regions and gene bodies (1) and appears to correlate with an expansion of the DNA methyltransferase and methyl-cytosine binding domain (MBD) families (2).

DNA methylation represents an abundant epigenetic mark in vertebrates, while many invertebrates lack detectable cytosine methylation (1). Notably, this is the case for several invertebrate organisms, including several invertebrate model organisms (e.g., yeast, fruit fly, and worm) (3). Most tested invertebrates, though, show mosaic methylation. But, the sea squirt *Ciona intestinalis* represents the only invertebrate for

which a detailed map of this pattern has been determined (4). Nevertheless, research continues to demonstrate the pervasiveness of cytosine methylation as an evolutionarily ancient genomic regulatory mechanism. Further, given that DNA methylation is a mechanism for regulating endogenous gene expression and possibly reducing transcriptional noise, research suggests that a reduction in unnecessary gene expression may have permitted the increase in gene number and in complexity that characterizes animals (5).

Sponges represent the most ancient and primitive animal (6) and share with fungi a common evolutionary history in that their last common ancestor was likely a flagellated protist similar to the choanoflagellate (7). Despite their distinct morphological separation from other animals, studies of sponges have shed light on the molecular evolution of animals. Like higher animals, the sponge genome harbors genes crucial for growth and differentiation, cell-cycle regulation, cell-cell adhesion, and self versus non-self recognition (8). Sponges also share in common with animals numerous transcription factor, sensory transduction, and cell adhesion genes (9).

However, little study of Poriferan epigenetic regulation has been carried out. Although the recently sequenced *Amphimedon queenslandica* genome highlights numerous molecular framework similarities between sponges and other animals, much is still unknown regarding the epigenetic machinery in place at the dawn of metazoans. Given the phylogenetic position of sponges at the base of the animal kingdom, elucidating their epigenetic machinery will better clarify epigenetic regulation carried out by the last common ancestor to all animals. This information will tell us more about the origins of the epigenetic machinery present in vertebrates. Characterizing these features

throughout animal developmental history could be highly useful regarding study of human epigenetics.

Vertebrates harbor a divergent family of methyl-cytosine binding domain (MBD) proteins critical for reading DNA methylation as an epigenetic mark. MBD family members include MeCP2 and MBD1-4. MBD1, MBD2, MBD3 are one sub-family: MBD1 and MBD2 prefer methylated DNA while MBD3 does not, at least not to the same extent. MBD4 and MeCP2 represent a second, more recent sub-family containing an insertion that leads to a longer alpha helix and larger hydrophobic core. MeCP2 is important for neurologic development while MBD4, through its associated glycosylase domain, repairs TpG/CpG mismatches (4).

MBD2 can be found throughout the animal kingdom and its structural components primarily consist of a methyl-cytosine binding domain followed by a region implicated in binding the RbAp46/48 homolog p55 (p55-binding region; p55BR) (10) and a C-terminal coiled-coil (CC) region. MBD2 targets methylated DNA with high affinity through highly conserved DNA contacting amino acid residues located within its methyl-cytosine binding domain.

MBD2 associates the ATP-dependent nucleosome remodeling and deacetylase complex (NuRD) (11). Six core proteins comprise the complex and constitute its ability regulate gene expression at the chromatin level: MBD2 or MBD3; histone deactylase (HDAC) 1 or HDAC2; chromatin remodeling enzyme Mi-2α (CHD3) or Mi-2β (CHD4); p66α (GATAD2A) or p66β (GATAD2B); metastasis-associated (MTA) protein (i.e., MTA1, MTA2, or MTA3); retinoblastoma-binding protein (RBBP; also known as RBAP48) 4 or RBBP7 (RBAP46). The MBD protein targets the complex to DNA while

HDAC1 or HDAC2 enzymatically removes acetyl groups from histone tails. The Mi-2 subunit carries out chromatin remodeling in part by repositioning nucleosomes. The p66α, p66β, RBBP4, and RBBP7 subunits are likely structural subunits of the complex that interact with histone tails. The p66α/β proteins contain two highly conserved regions: an N-terminal coiled-coil domain (CR1) and a C-terminal GATA-like zinc finger domain (CR2) (12, 13). The MTA protein is thought to localize the complex to different cell-specific targets by associating with transcription factors (14).

The coiled-coil region of human MBD2 binds the p66α coiled-coil domain (conserved region 1- CR1) forming a high affinity heterodimeric complex that recruits Mi-2α/β and correlates with *in vivo* methylation dependent transcriptional repression (15). This unique coiled-coil interaction results in a heterodimeric complex that forms from anti-parallel association between two peptides that are largely monomeric in isolation. Helical content and specific electrostatic interactions between charged residues on the individual coiled coil domains drive the high affinity binding observed (16). This suggests that isoforms of MBD2 and p66α exhibiting a reduction in helical content will bind p66α with reduced affinity.

The MBD2 p55BR lies between the methyl-cytosine binding domain (MBD) and the coiled-coil domains of MBD2. The first two-thirds of the region recruits the NuRD core components RbAp46/48, HDAC1/2 and MTA1/2. Furthermore, two conserved residues Arg286 and Leu287, establish critical contact points for complex formation (Desai, MA, unpublished data). This region has been determined to display an evolutionarily conserved, positive isoelectric point along with a largely unstructured architecture in isolation. Despite its unstructured architecture, the p55BR enhances the

affinity of MBD2 to methylated DNA by approximately 100-fold. (Desai, MA, unpublished data).

NuRD complex components have previously been identified in organisms throughout the metazoan line including plants and worms, and more recently, in the sponges *Amphimedon queenslandica* (9) and *Ephydatia muelleri* (unpublished data, Pohlman D.). Whether or not the gene products identified in these two sponge species form an intact and/or functional NuRD complex remains unclear. Given the characteristics of the human MBD2, and if methylation patterns are intrinsically different between vertebrates and invertebrates, we initially asked the following questions: (1) Does the MBD2 complex still target methylated DNA? (2) Does MBD2 recruit NuRD components in the same manner? (3) What are the functional consequences of MBD2 knockdown?

To answer these questions, we cloned, expressed, purified, and determined the methylation selectivity of *Ephydatia* MBD2 MBD by surface plasmon resonance (SPR). We also cloned, expressed, and purified the coiled-coil domains from *Ephydatia* MBD2 and p66α and assessed complex formation using analytical ultracentrifugation (AUC) and isothermal titration calorimetry (ITC). We assessed the temperature-dependent helical properties of *Ephydatia* MBD2 and p66α in isolation and in complex using circular dichroism (CD). We further cloned, expressed, and purified the full-length *Ephydatia* p55BR along with a shorter version of the protein consisting of the first two-thirds of the protein. RNAi-mediated knockdown of *Ephydatia* MBD2 and observation of subsequent phenotypic effects was carried out by a collaborator.

Here, we show that *Ephydatia* MBD2 clearly prefers methylated DNA with a

binding affinity close to that of human and chicken. However, the binding constant of the sponge MBD2-p66α complex is approximately 1000-fold weaker than the human complex. ITC of *E.m.*-*H.s.* complexes (either *E.m.* p66α:*H.s.* MBD2 or *E.m.* MBD2:*H.s.* p66α) indicate higher affinity binding. We performed cross-species pull-down assays involving the *Ephydatia* MBD2 p55BR and found that the protein recruits human MTA2 but not human HDAC1/2 or RbAp 46/48. RNAi-mediated knockdown of *Ephydatia* MBD2 results in a low-growth phenotype marked by failure to form structures associated with early development in sponges. These data suggest that sponges likely have the capacity to carry out sophisticated epigenetic processes, including targeted chromatin remodeling, that likely predate sponge emergence.

3.3 EXPERIMENTAL PROCEDURES

3.3.1 Purification of proteins and DNA:

The coiled-coil regions of human MBD2b (amino acids 211-244), p66α (amino acids 137-178) and *Ephydatia muelleri* MBD2 (amino acids 182-212) and p66α (amino acids 58-101) were cloned and expressed with a hexahistidine tag and as thioredoxin fusion proteins in a modified pET32a vector. The expression vectors were transformed into the Rosetta BL21 (DE3- Invitrogen) *E. coli* strain, grown in Luria Bertani medium at 37°C and induced with 1 mM isopropyl-β-d-thiogalactopyranoside at an A_{600} ~ 0.8. The bacteria were harvested after 2 hours of induction and lysed ultrasonically in the presence of B-PER reagent (Thermo Scientific). The soluble fraction was passed over a nickel-sepharose column with the tagged protein eluted by a step

gradient of imidazole and further purified by gel filtration over a Superdex-75 column (GE Healthcare). The thioredoxin fusion proteins were used directly for analytical ultracentrifugation (AUC) and isothermal titration calorimetry (ITC) studies. For circular dichroism (CD) studies, clones were modified to incorporate a tyrosine residue just after the thrombin cleavage site (for quantification of the isolated peptide by UV measurement) and were expressed in a similar manner. After purification over a nickel-sepharose column, the thioredoxin tag was separated from the rest of the peptide by thrombin digest and isolated by gel filtration chromatography over a Superdex-75 column (GE Healthcare). Specific mutations were introduced using the QuikChange Lightning® site-directed mutagenesis kit (Stratagene) following the manufacturers protocol. The final concentrations of all protein samples were determined by UV absorbance at 280 nm. The 17-bp complementary oligonucleotides (Table 2 of the previous chapter) were purchased (Integrated DNA Technologies), annealed and purified as previously described. The sequences were derived from the $p16^{INK4a}$ promoter known to be a native target sequence for MBD2, as mentioned in the previous chapter.

Two *E.m.* MBD2 p55BR constructs were cloned into the pCMVTag2B (Stratagene) vector in frame with an N-terminal Flag-tag sequence: full-length p55BR (amino acids 90-181) and short (amino acids 90-181). These constructs were used in the co-immunoprecipitation procedure discussed below.

3.3.2 Binding affinity:

Binding affinities were determined by surface plasmon resonance analysis on a Biacore T100 system (GE Healthcare) as described previously. The binding affinity was determined from steady state analysis of the SPR relative response at varying

concentrations of protein. As previously shown, the maximum steady state response (R_{max}) in SPR depends on stoichiometry (n) of binding. Prior to fitting, the steady state response at each protein concentration (R^A) was normalized (R_{norm}) to the total DNA immobilized (Rl) and molecular weights of the DNA and protein (MW_L and MW_A, respectively).

$$R_{norm} = \frac{R_A}{R_L \bullet \left(\frac{MW_A}{MW_L} \right)} \qquad (1)$$

Final data analysis, plotting, and curve fitting were performed with pro Fit software (QuantumSoft).

3.3.3 Analytical ultracentrifugation:

Sedimentation velocity experiments were carried out using Beckman Optima XL-I analytical ultracentrifuge (Beckman Coulter Inc.) equipped with a four and eight-position AN-60Ti rotor. Sedimentation was performed at 40,000 rpm, 20°C, under physiological buffer conditions (20 mM Tris pH 8.0, 150 mM NaCl). Sedimentation profiles were recorded using UV absorption (280 nm) and interference scanning optics. The partial specific volume (V) of the sample, density (ρ) and viscosity (η) of the buffer were calculated using the SED NTERP program (25). Data was fit using a continuous size distribution (c(s)) and the effective molecular weight determined from the resulting sedimentation coefficients with the SEDFIT software (26).

3.3.4 Isothermal titration calorimetry:

Protein samples were buffer exchanged into (50 mM Hepes pH 7.4, 150 mM NaCl) and binding analyzed with an iTC200 Microcalorimeter (GE Healthcare). A total of 24 injections (1.5 μL each) of the p66α coiled-coil (100 μM) were injected into MBD2 and homologues (10 μM, 298 K, stir speed of 400 rpm, 120 seconds time delay between injections). The resulting isotherms were auto adjusted for baseline and fit to a one-site binding model using Origin 7.0 software to determine binding constant (K_D) and enthalpy (ΔH) while the Gibbs free energy (ΔG) and entropy (ΔS) of binding were calculated according to Equation 1.1,

$$-RTln(K) = \Delta G = \Delta H - T\Delta S \text{ (Equation 1.1),}$$

where T is the temperature in Kelvin and R is the gas constant.

3.3.5 Circular dichroism:

CD spectra were collected on purified peptide samples (~33 mg/mL total protein in 10mM sodium phosphate, pH 6.5) with a JASCO J-715 CD spectrometer (JASCO Corp) at 293 K, with a 1-cm path length, scanning from 190-260 nm with 0.5 nm interval at a scanning speed of 50 nm/min. CD spectra were normalized to give molar ellipticity values(θ) in degrees·cm^2·dmol-1residue.

Helical content for each peptide was calculated from the ratio of the observed θ222nm to the expected θ222nm for 100% helix as given by 40,000 × [(n − 4)/n], where n is the number of residues. *Thermal denaturation was followed at θ222nm* from 277-368K at 1 K intervals with a heating rate of 1 K/min. The data were fit to a simple two state thermodynamic model of unfolding.

Helical content prediction

The expected helical content for each peptide was calculated using the AGADIR (27) algorithm with the N- and C-termini 'free', at 293 K, ionic strength of 0.02, and pH 6.5 to closely match the experimental conditions for CD. The predicted helical content was used to help design amino acid changes that stabilize helix formation.

3.3.6 Co-immunoprecipitation studies:

Various MBD2 constructs were cloned into the pCMVTag2B (Stratagene) vector in frame with an N-terminal Flag-tag sequence. HEK 293T cells were transfected with the constructs (18 ug plasmid DNA) by calcium phosphate precipitation method [17] and harvested after 48 hours. Cells were lysed and immunoprecipitated with anti-Flag M2 antibody (Sigma) and mouse IgG (Santa Cruz) controls according to the Sigma Flag-IPT kit protocol (Sigma-Aldrich, Inc., St. Louis, MO). The precipitated proteins were then analyzed for different components of the MBD2/ NuRD complex by western blot using antibodies against RbAp48 (Abcam), HDAC2 (Millipore) and MTA2 (Santa Cruz).

3.3.7 RNAi studies:

The following MBD2 RNAi procedure carried out in Ephydatia muelleri was provided by April Hill, Ph.D., Department of Biology, University of Richmond, Richmond, Virginia.

Ephydatia muelleri gemmules were grown following the protocol in Rivera et al. 2011, with 2-3 gemmules sharing a single well. After the gemmules attached, 10,000 ng of dsRNA was mixed with 1 mL of Strekal's media in 3 of the six wells in which the

sponges grew (Strekal and McDiffett 1974). After 24h, media was changed, and the same 3 wells received another 10,000ng of dsRNA. 48h after the original treatment, gemmules were scored for growth, size, number of oscula, and canal structure. They were photographed on a stereomicroscope.

dsRNA Synthesis:

The methyl binding domain of MBD2/3 was amplified from *Ephydatia* cDNA using these primers: F: 5' – CGACTCACTATAGGGTTGCGGGTGAAGGGCTTG – 3', R: 5' – CGACTCACTATAGGGGCCATCGCCAGGGGAAC – 3'. The primers contain partial T7 sites that flank the ~204bp amplicon. The full T7 site was added in a subsequent PCR, using a complementary primer that matched the partial T7 site. Both reactions were run in the thermocycler at: 94°C for 3 min, 30 cycles of 94°C for 30s, 57°C for 30s, 72°C for 1 min, followed by a 5 minute final extension at 72°C. PCR products were again run on a gel and the correct bands excised and purified. This product was used to make double-stranded RNA (dsRNA) using the T7 RiboMAXTM Express RNAi System (Promega). dsRNA was quantified using a NanoDrop.

3.4 RESULTS

3.4.1 Ephydatia muelleri MBD2 targets methylated DNA.

Alignment analysis indicates the *Ephydatia* and human MBD2 methyl-cytosine binding domains share high identity (73%) as seen in figure 3.2, but less similarity throughout the remainder of MBD2 (see the clustal analysis- figure 3.3). As can be observed in figures 3.2 and 3.3, the sponge sequence contains all amino acids known to

contact DNA and found in the loop connecting the central two beta strands of the beta sheet: R24, G27, A30, H32, D34, F36, R46. This, along with the overall high identity between *Ephydatia* and human MBD2 MBD, prompted us to surmise that the sponge MBD would target methylated-CpG sites with similarly high affinity. To assess the binding affinity of the *Ephydatia* MBD2 MBD for methylated and unmethylated sequences DNA steady-state analysis of surface plasmon resonance data was conducted, as described in the previous chapter. To allow for direct comparison of binding stoichiometry, we normalized the relative steady-state response to the amount of DNA coupled to the sensor chip such that the maximum steady-state response reflects the number of binding sites on the DNA. As can be seen in Fig. 3.4 and Table 3.1, both the human and *Ephydatia* MBD2 MBD bind mCpG DNA with high affinity and stoichiometry of approximately one (K_D = 0.19 ± 0.03 μM and 0.43 ± 0.09 μM, respectively and R_{max} = 1.2, for both human and *Ephydatia*). Both bind unmethylated DNA with much lower affinity and high stoichiometry (K_D = 82 ± 5 μM and 65 ± 2 μM and R_{max} = 7.3 and 7.9, respectively) indicative of a high degree of methylation selectivity. However, human MBD2 selects methylated DNA with a 430-fold higher affinity over unmethylated, whereas, *Ephydatia* MBD2 binds with a 150-fold higher affinity than unmethylated DNA.

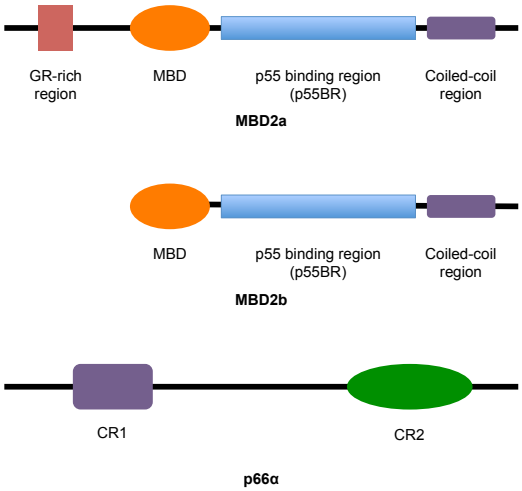


Figure 3.1: Domain organization of the *Homo sapiens* MBD2 and p66α. GR, glycine-arginine repeat region; MBD, methyl-binding domain; CR1 is a coiled-coil domain; CR2 includes a GATA zinc finger domain.

```
ALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDF  58  Human MBD2 MBD
 LP GWK+E V+RK+G SAGK+DVYYFSP GKKFRSKPQ+AR+LG+ VDL+ FDF
GLPSGWKREVVVRKNGQSAGKTDVYYFSPCGKKFRSKPQIARFLGDAVDLTCFDF  59  Ephydatia MBD2 MBD
```

Figure 3.2: Alignment of the *Homo sapiens* and *Ephydatia muelleri* MBD2 methyl-cytosine binding domains.

```
EmMBD2a         MSSSSPPEYETVRTELRVKGLPSGWKREVVVRKNGQSAGKTDVYYFSPCGKKFRSKPQIA 60
EmMBD2b         MSSSSPPEYETVRTELRVKGLPSGWKREVVVRKNGQSAGKTDVYYFSPCGKKFRSKPQIA 60
HsMBD2b         ---------------MDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLA 45
                               :   .**.***:* *:**.* ****:*******.*********:*

EmMBD2a         RFLGDAVDLTCFDFSRAGSPGDGTQRRRARDRNLGSPGGGGGGGGGAGGHHSHSNGGTSS 120
EmMBD2b         RFLGDAVDLTCFDFSRAGSPGDGTQRRRARDRN--------------------------- 93
HsMBD2b         RYLGNTVDLSSFDFRTGKMMPSKLQKNKQRLRN--------------------------- 78
                *:**::***:.***  .    .  *:.: * **

EmMBD2a         AGNIVTRRSDHKPSSGSKPLSTNPLRP----SGPVRRTCGVIKLPVIWVAPPNNEQVRDE 176
EmMBD2b         -----------------KPLSTNPLRP----SGPVRRTCGVIKLPVIWVAPPNNEQVRDE 132
HsMBD2b         -----------------DPLNQNKGKPDLNTTLPIRQTASIFKQPVTKVTNHPSNKVKSD 121
                               .**. *  :*    : *:*:*..::* **  *:   .::*:.:

EmMBD2a         CVILDQKRDVAVHVVVQTHWERRLGGFKPCDHATGSDIAFHRVPNGTVDGSATKKPAASQ 236
EmMBD2b         CVILDQKRDVAVHVVVQTHWERRLGGFKPCDHATGSDIAFHRVPNGT------------- 179
HsMBD2b         PQRMNEQP-------RQLFWEKRLQGLS-ASDVTEQIIKTMELPKGLQG----------- 162
                   ::::         * .**:** *:. ....* . *   .:*:*

EmMBD2a         QTPAASASSKTQSSVLSHLPTPFLTSASPVSSLASQSPRSTNAPAGPPVHSGTDTLSANG 296
EmMBD2b         ------------------------------------------------------------
HsMBD2b         -----VGPGSNDETLLSAVASALHTSSAPITG---------------------------- 189


EmMBD2a         PSASQQQTHSSTHSLVSNGSSAQTVELSLPLVTESVVRAQEERVRLIRQQLLAAQSISS- 355
EmMBD2b         -------------------------LSLPLVTESVVRAQEERVRLIRQQLLAAQSISS- 212
HsMBD2b         ------QVSAAVEKNPAVWLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEEALMADIL 243
                                     .  :**:. :* *****: :*::*  *   .

EmMBD2a         --------------------
EmMBD2b         --------------------
HsMBD2b         SRAADTEEMDIEMDSGDEA 262
```

Figure 3.3: Clustal alignment of the *Homo sapiens* MBD2 versus *Ephydatia muelleri* MBD2a and MBD2b.

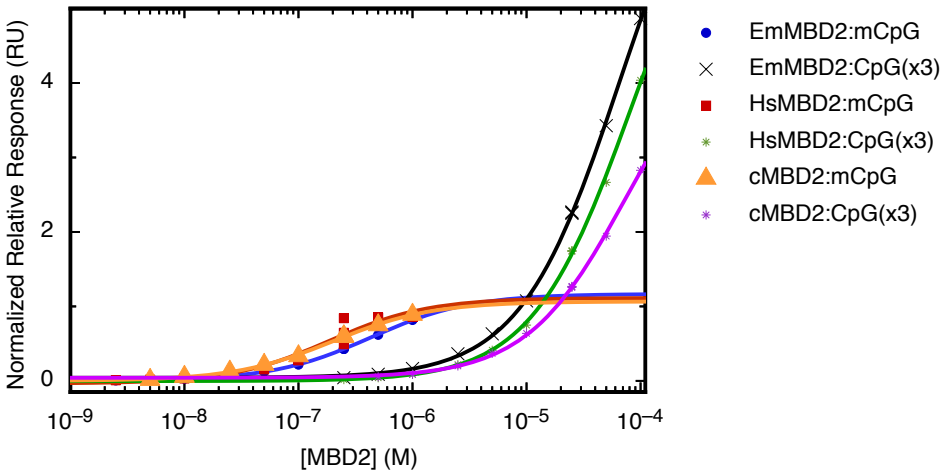


Figure 3.4: Ephydatia MBD2 selectively targets methylated DNA. Steady state surface plasmon resonance measurements are shown for MBD2 binding to immobilized double stranded methylated or unmethylated oligonucleotides. The steady-state response was normalized to the amount of DNA immobilized such that the maximum response reflects the stoichiometry of binding (n), as described in the previous chapter.

Protein	DNA	Kd (μM) ± SE	R_{max} *	Chi^2 ($x10^{-3}$)
HsMBD2	mCpG	0.19 ± 0.03	1.2	87
HsMBD2	CpG(x3)	82 ± 5	7.3	17
cMBD2	mCpG	0.20 ± 0.01	1.0	0.90
cMBD2	CpG(x3)	74 ± 4	4.8	3.8
EmMBD2	mCpG	0.43 ± 0.09	1.2	0.10
EmMBD2	CpG (x3)	65 ± 2	7.9	8.1

*Normalized such that R_{max} reflects stoichiometry (n)

Table 3.1: Binding affinity. The dissociation constant (K_D), R_{max}, and Chi^2 are given for different protein and DNA complexes as determined by steady state analysis of surface plasmon resonance studies.

3.4.2 Key amino acid contact residues involved in high-affinity binding between the human MBD2 and p66α coiled-coil regions are found in Ephydatia muelleri.

The human and *Ephydatia* MBD2 CC regions share 46% identity, whereas the p66α peptides share only 37% identity. Yet, the key amino acid residues involved in the human MBD2 and p66*α* coiled-coil complex are largely conserved in *Ephydatia* (Figures 3.1 and 3.5), suggesting that, like the human CC regions, binding is largely driven by key hydrophobic and electrostatic contacts, and the helical properties of the peptides.

The coiled-coil typically consists of two or more α-helices wrapped around one another. Coiled-coils are characterized by amino acid sequences forming a heptad repeat pattern, with positions designated *abcdefg*; dimerization interfaces form from hydrophobic side chains at the *a* and *d* positions. In this sequence, the first and fourth residues are hydrophobic, and residues in the fifth and seventh position are primarily charged or polar (18). Coiled-coil segments may be unfolded in monomeric form and then fold during interaction with a binding partner (19). A two-state transition often describes the folding and unfolding of coiled coils, i.e., unfolded peptide monomers cooperatively fold to form coiled-coil dimers or oligomers (18).

Parallel and antiparallel coiled-coil dimerization results largely from the interdigitation and burial of hydrophobic side chains of one helix into its partner. The human MBD2/p66*α* CC interaction represents an anti-parallel orientation with intermolecular, vertical contacts between branched hydrophobic residues. The vertical interactions form between residues in the *a* position of the heptad repeat of one helix and the *d'* of the other, and the *a'* and *d* positions. Triplet repeats of branched hydrophobic residues placed in the *a'-a-a'* positions further promote heterodimerization (16). The

position of hydrophobic residues in human MBD2 and p66*α* represents an arrangement found to strongly favor heterodimeric, antiparallel coiled-coil formation (16, 20). More specifically, hydrophobic interdigitation between the two coiled-coils presents as *R*I*L*V*LL*I (p66*α* residues are in italics) (16).

As mentioned previously, *Ephydatia* MBD2 p66*α* CC regions share high identity with their human counterparts, regarding the amino acid residues involved in the coiled-coil binding interface. *E.m.* MBD2 presents a VVL sequence rather than IVL, while *Ephydatia* p66*α* offers an RLLV the *a-a'* interdigitated sequence between the *Ephydatia* peptides is likely represented by *R*V*L*V*LL*V. Since the side chain of valine is shorter than leucine or isoleucine, the presence of valine-227 in human MBD2 may allow closer proximity, and therefore tighter packing, of the surrounding side chains at the interface. Conversely, the additional valines present at the *E.m.* interface may prevent tighter packing, as the valine side chain may be unable to reach into the binding pockets as well as the isoleucine or leucine residues present in the human peptides.

When the glutamate at position 224 of human MBD2 was mutated to a glycine, decreases were observed in both helicity and binding to p66*α*, showing that pre-formed helical content is necessary for binding; this residue is not found on the binding surface between the two CCs (16). The coiled-coil regions of *Ephydatia* MBD2 and p66*α* both have several amino acids that disrupt alpha-helicity (Figure 3.2). Since the high affinity interaction between human MBD2 and p66*α* was shown to be strongly influenced by the helical propensity of both prior to binding, the *Ephydatia* orthologs are unlikely to display the same high helical content and, therefore, the same degree of binding as the human peptides.

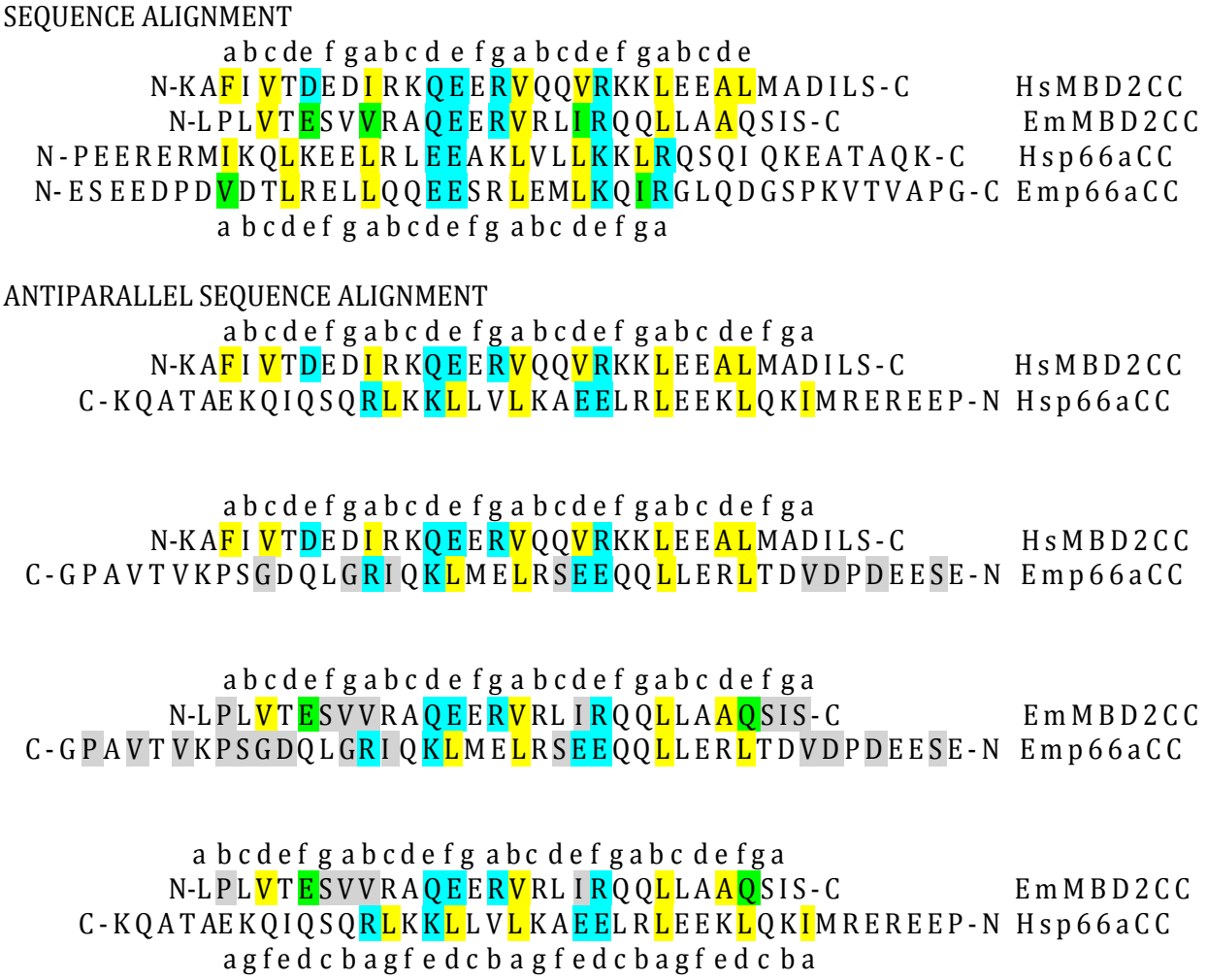


Figure 3.5: Human and *Ephydatia MBD2* and *p66α* coiled-coil sequence alignments. Top, a sequence alignment of the coiled-coil (cc) domains from human and *Ephydatia* p66*α* and MBD2 is shown with key hydrophobic (yellow) and ionic/polar (cyan) contact residues highlighted and the heptad repeat (a-g) indicated above and below the amino acid sequences. Bottom, an anti-parallel sequence alignment showing the position of contact residues between the human and sponge proteins, with amino acid residues known to disrupt alpha-helicity highlighted (grey) for Ephydatia.

3.4.3 Ephydatia MBD2 and p66α coiled-coil regions remain largely monomeric in solution.

The human MBD2 coiled-coil region does not form homo-oligomers as often observed with coiled-coils. Indeed, the protein remains monomeric in isolation up to a concentration of 300 μM. The human p66*α* coiled-coil, however, forms homo-dimers in concentrations above 50 μM, though the monomer remains the dominant form (15). Preliminary size exclusion chromatography (SEC) experiments suggested weak binding between the *Ephydatia* MBD2 MBD and p66*α* coiled-coil regions. We then performed sedimentation velocity AUC studies so as to assess the homo-dimerization versus homo-oligomerization potential of each peptide in isolation and the hetero-dimerization potential in complex. These results indicate that *Ephydatia* MBD2 coiled-coil also remains monomeric in isolation up to a concentration of 100 μM (Figure 3.6). Though the *Ephydatia* p66*α* coiled-coil region largely exists as a monomer at a 100 μM concentration, it does exhibit a tendency to homo-dimerize. The sedimentation velocity study also indicates that when the two proteins are combined at final concentrations of 100 μM, monomers, heterodimers, and hetero-oligomers exist in solution. Given that the binding constant between the two sponge coiled-coils is 25 μM, a concentration at only four times the K_D (100 μM) of a weakly binding complex, is likely to show multiple species.

The weak binding suggested by SEC and AUC, along with the low helical propensity predicted by Agadir (see results below), prompted us to turn our attention to CD and ITC analyses to further explore the helical and binding properties of the Ephydatia coiled-coil proteins.

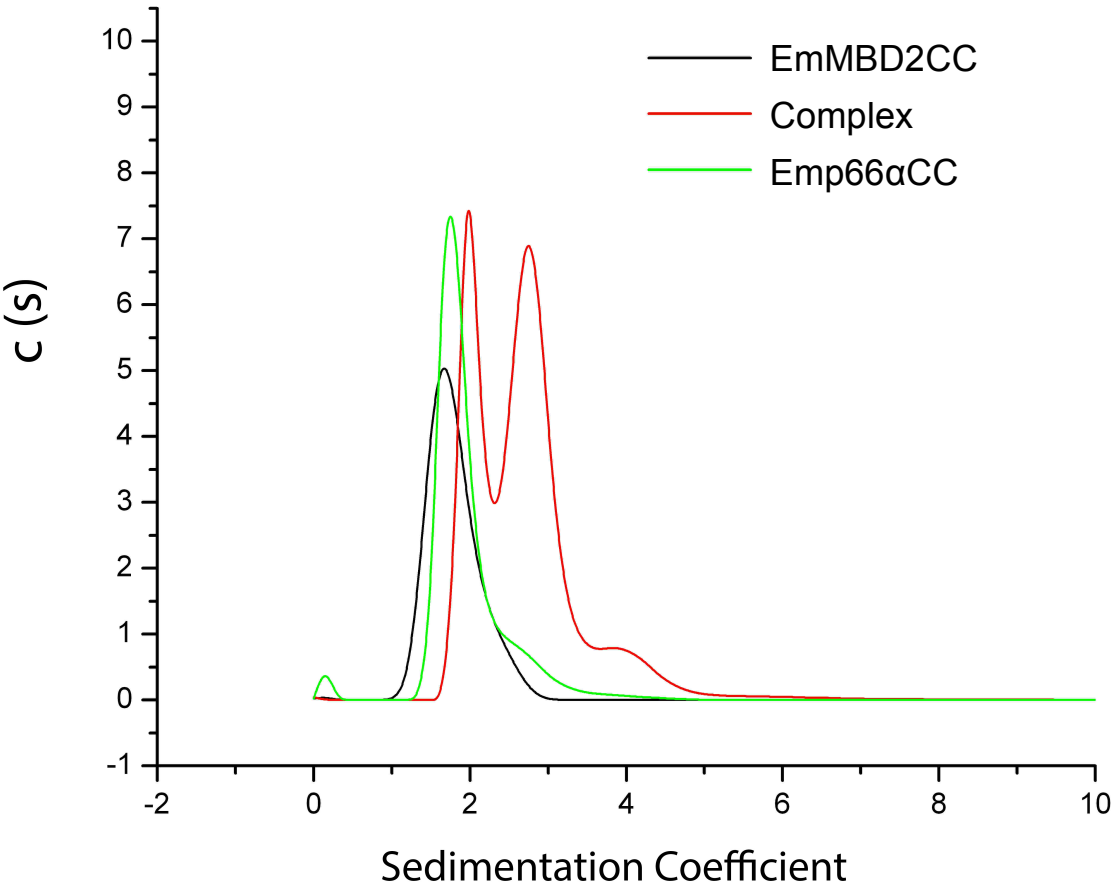


Figure 3.6: The *Ephydatia* coiled-coil domains remain largely monomeric in isolation. Analytical ultracentrifugation analysis was performed on the individual coiled-coil domains and the sedimentation velocity fit using a continuous size distribution (c(s)). The results are shown for 100 uM concentrations of MBD2 and p66α coiled-coil domains.

3.4.4 Ephydatia MBD2 and p66α display lower helical propensity than their human counterparts.

Agadir is a prediction algorithm that predicts helical propensity of monomeric peptides in solution, and is based on the helix/coil transition theory (21). The algorithm predicts that the *Ephydatia* MBD2 and p66*α* coiled-coil domains do not have the same tendency to form *α*-helices in isolation as with the human orthologs (Table 3.2). CD analyses were performed on the isolated domains, so as to determine the relative helical

content of the coiled-coil regions in isolation in comparison to that predicted by Agadir.

Human MBD2 (25%) and p66*α* (66%) are more helical than *Ephydatia* MBD2 (5%) and p66*α* (20%). The thermal stability of the different coiled-coil complexes was determined by following molar ellipticity at 222 nm ($\theta_{222\ nm}$) as a function of temperature. No cooperative transition was observed during the melt of the *Ephydatia* coiled-coils while the human variants do experience this phenomenon. High affinity binding between human MBD2 and p66*α* depends on the helical propensity of their coiled-coils (16). Since numerous amino acids known to disrupt alpha-helical structure are present in both Ephydatia CC regions, we therefore predicted a lower binding constant for the *Ephydatia* coiled-coil interaction.

The human complex melts at 338K while the sponge complex does not exhibit a cooperative phase transition associated with melting (figure 3.7, bottom). However, the complex could only be accurately measured at 13uM, which is below the 25 uM binding constant; only a small fraction would be bound at this concentration. The predicted versus calculated helicities differed significantly for the *Ephydatia* coiled-coils. Limitations for the Agadir algorithm exist that cause the program to over or underestimate the helical content of peptides; these factors may apply in this scenario. For example, the algorithm contains no mode for assessing side-chain interactions. The algorithm also assumes that there is no energy coupling, other than electrostatic, between residues in the random-coil state, although interactions may also result in the formation of beta-turns in solution; the algorithm may predict a random-coil. Further, peptides may aggregate in solution thereby altering helical propensity. This is another case that Agadir is unlikely to incorporate into the prediction. Also, the presence of the tyrosine residue at the beginning of the amino

acid sequence likely disrupts the helicity of the peptide; side-chain interactions with the peptide helix may alter both the geometry and CD signal. CD analysis of polyalanine-based peptides with an N-terminal tyrosine residue located at either end of the chain sometimes associates with a positive band at 222 nm (22). Thus, a discrepancy emerges between the predicted and calculated (i.e., an over-estimation of helicity by Agadir).

	Helical propensity in isolation	
Coiled-coil domain	Predicted[a]	Calculated[b]
HsMBD2	40	25
Hsp66α	55	66
EmMBD2	26	5
Emp66α	8	20

[a] Based on the AGADIR algorithm
[b] Based on CD measurements

Table 3.2: Helical content of the Ephydatia MBD2 and p66α coiled-coils. Percent helicity as predicted by AGADIR and calculated from the circular dichroism molar ellipticity at 222 nm ($\theta_{222\ nm}$) is given for the *Ephydatia muelleri* and *Homo sapiens* coiled-coil domains.

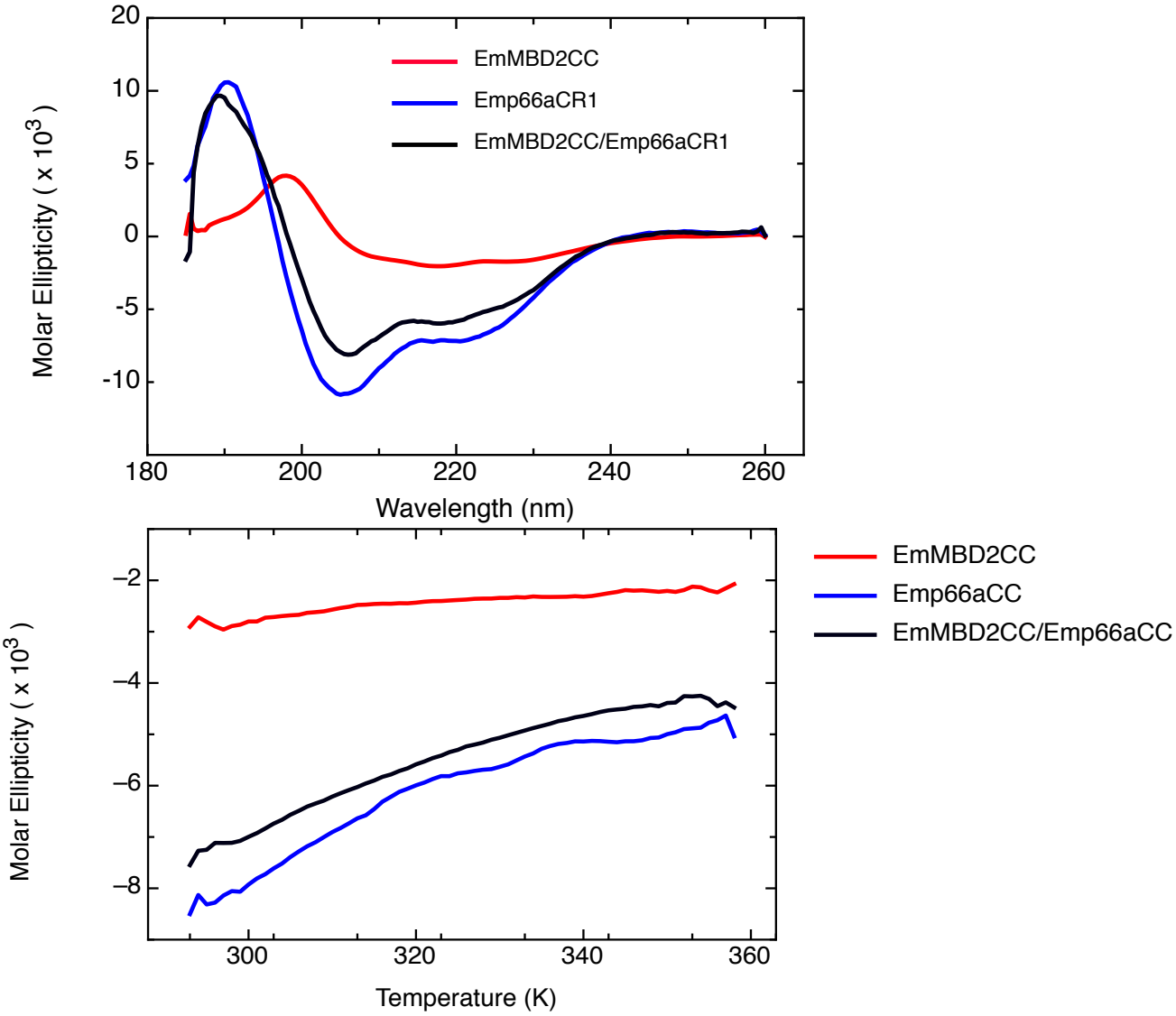


Figure 3.7: Circular dichroism spectra of Ephydatia coiled-coils in isolation and in complex. Top: circular dichroism spectra. Bottom: a thermal melt indicating the temperature dependence of the molar ellipticity at 222 nm from 293 to 358K.

3.4.5 Ephydatia MBD2 and *p66α coiled-coil regions bind weakly, but bind to human p66α and MBD2 coiled-coils, respectively, with higher affinity.*

Preliminary SEC evidence suggested that an *Ephydatia muelleri* (*E.m.*) MBD2-p66*α* complex binds very weakly (data not shown), converse to the tight binding experience by the coiled-coils in the human (*H.s.*) complex. We set out to more accurately determine the binding constant for the *E.m.-E.m.* complex. Our lab previously determined that the high affinity of the human coiled-coils results from conserved hydrophobic and electrostatic interactions between the coiled-coils (CC) and their helical content. Since the *Ephydatia* coiled-coil peptides largely contain the same interacting residues, and yet likely experience weak binding as a result of predicted reduced helicity, we were curious if the respective *Ephydatia* coiled-coil peptides would form higher affinity complexes when combined with the human coiled-coil regions (i.e., *E.m.* MBD2/*H.s.* p66α, *E.m.* p66α/*H.s*. MBD2).

Preliminary SEC data suggested that the *E.m.-H.s.* complexes bind with higher affinity (data not shown). ITC was then performed using thioredoxin fusion constructs of the coiled-coil domains. Exothermic heat was generated with each injection in all experiments. The binding isotherms (Figures 3.8 and 3.9), and the measured binding affinity (K_D), free energy (ΔG), enthalpy (ΔH), and entropy ($-\Delta T\Delta S$) for each complex (Table 3.3) show that indeed the *Ephydatia* peptides bind with higher affinity when in complex with the human (*H.s.* p66αCR1/*E.m.* MBD2CC: K_D- 390 ± 30 nM; *E.m.* p66αCR1/*H.s.* MBD2CC: K_D- 1.3 ± 0.11 μM) as compared with the *Ephydatia* only complex (24 ± 0.12 μM). Each complex binds with a stoichiometry of approximately 1:1 (n ranges from 0.99 to 1.4, Table 1) consistent with heterodimer formation.

Of the binding interactions tested, *E.m.* p66α for *E.m.* MBD2 complex formation reflects the most unfavorable change in entropy upon binding (TΔS: -1.097 kcal/mol) despite a favorable change in enthalpy (ΔH -7.405 ± 0.1001 kcal/mol) supporting the notion that pre-formed helicity is an important determinant of binding affinity. The ITC results further establish a hierarchy regarding favorability of complex formation: *H.s.*-*H.s.*>*E.m.* MBD2-*H.s.* p66α >*H.s.* MBD2-*E.m.* p66α >*E.m.*-*E.m.* (highest to lowest favorability).

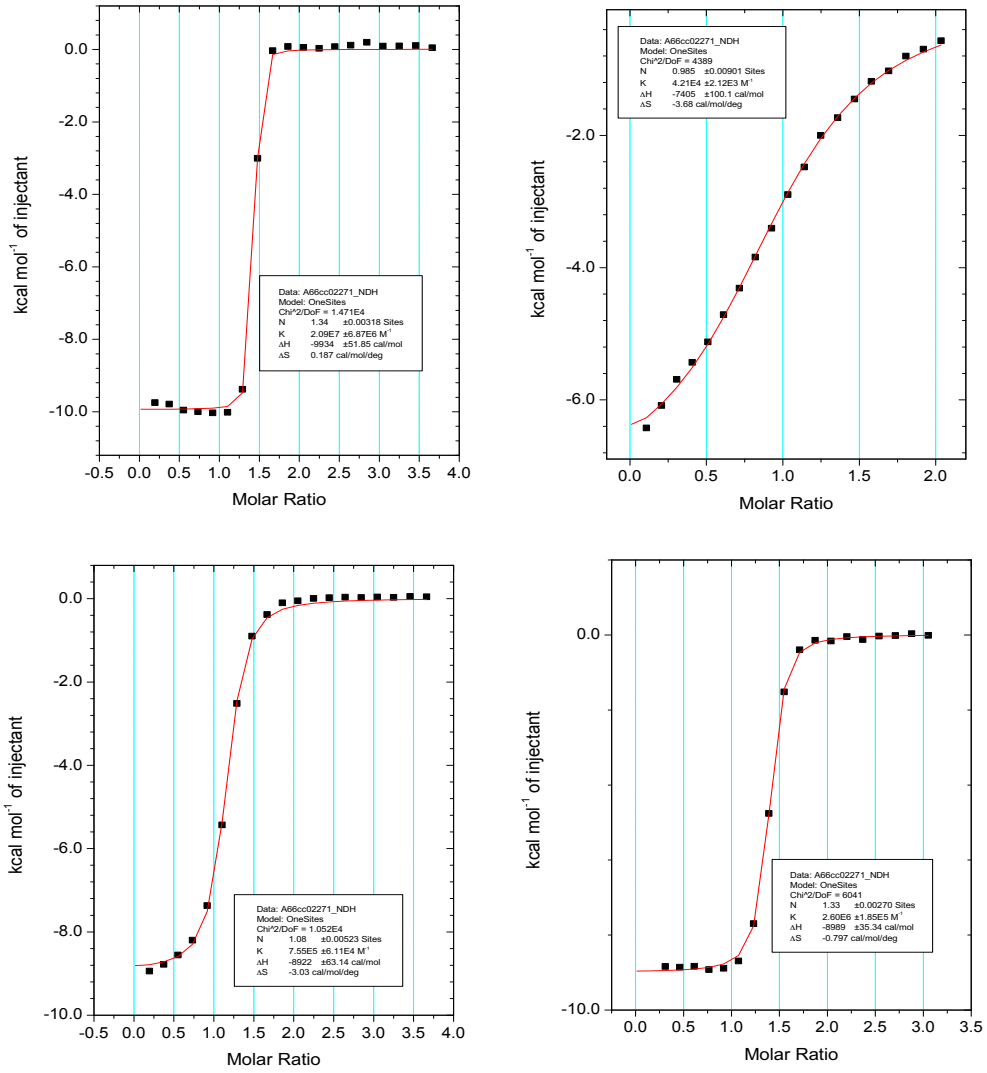


Figure 3.8: Binding analysis of *Ephydatia* and human MBD2 and p66α coiled-coil interactions. Isothermal titration calorimetry studies indicating the resulting fit and associated data are shown for human and *Ephydatia* MBD2 and p66α coiled-coil domains binding interactions. Top left: human MBD2 and p66α. Top right: *Ephydatia* MBD2 and p66α. Bottom left: Human MBD2 and *Ephydatia* p66α. Bottom right: Human p66α and *Ephydatia* MBD2.

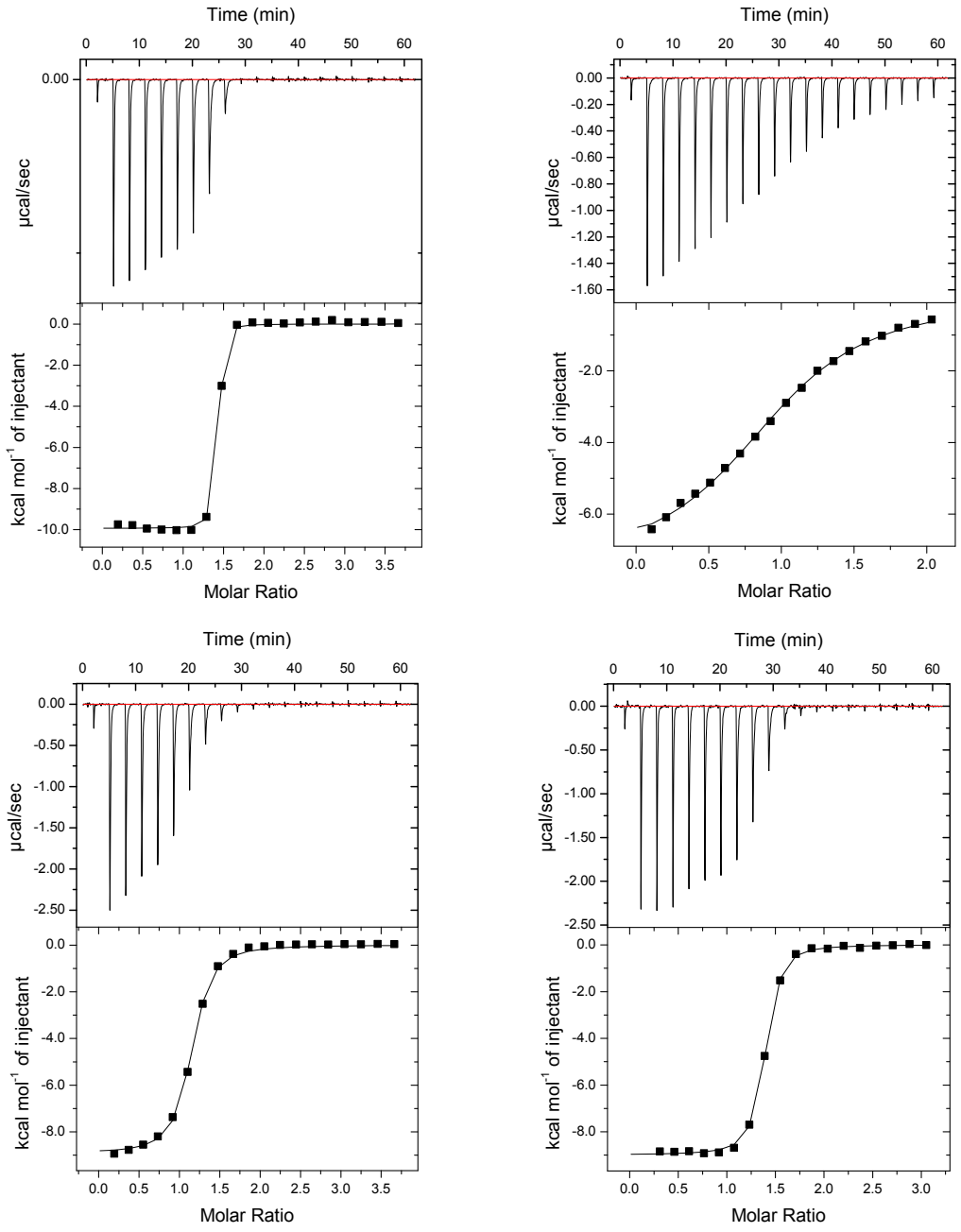


Figure 3.9: Binding analysis of *Ephydatia* and human MBD2 and p66α coiled-coil interactions. Isothermal titration calorimetry studies indicating the experimental data (top panel) and resulting fit (bottom panel) are shown for human and *Ephydatia* MBD2 and p66α coiled-coil domains binding interactions. Top left: human MBD2 and p66α. Top right: *Ephydatia* MBD2 and p66α. Bottom left: Human MBD2 and *Ephydatia* p66α. Bottom right: Human p66α and *Ephydatia* MBD2.

	K_D (nM)	n	ΔH (cal/mol)	TΔS (cal/mol)	ΔG (cal/mol)
Hsp66αCR1/HsMBD2CC	48 ± 7	1.3	-9,934 ±127.6	-55.7	-9,878
Hsp66αCR1/EmMBD2CC	390 ± 30	1.4	-8,989 ±103.6	-238	-8,751
Emp66αCR1/HsMBD2CC	1,300 ± 110	1.3	-8,922 ±63.14	-903	-8,019
Emp66αCR1/EmMBD2CC	24,000 ± 1,200	0.99	-7,405 ±100.1	-1,100	-6315

Table 3.3: Binding affinity analyses of *Ephydatia* and human MBD2 and p66α coiled-coil interactions. The dissociation constant (K_D), change in enthalpy (ΔH), entropy (-TΔS), Gibbs free energy (ΔG), and apparent stoichiometry (n) derived from isothermal titration calorimetry studies are given for the coiled-coil complexes between human and *Ephydatia* p66α and MBD2.

3.4.6 The Ephydatia p55 binding region does not recruit human NuRD core components.

The region between the MBD2 MBD and coiled-coil domain enhances the affinity of MBD2 for DNA and behaves as an intrinsically disordered region in isolation, in the context of the full-length protein, and when bound to DNA (Desai, M., et al., unpublished data). The first and second regions of the human MBD2 p55BR (amino acids 240-316) are sufficient to recruit RbAp48, HDAC2, and MTA2, although showing a somewhat weaker interaction with MTA2 as compared to the full-length p55BR (Desai, M., et al., unpublished data). The human and sponge MBD2 p55BR share 30% identity and nearly identical isoelectric points (pI approximately 10). Without a reliable means of assessing the capacity of *E. muelleri* MBD2 p55BR to recruit NuRD components in sponge cells, we investigated whether or not the *E. muelleri* p55BR interacts with the core components of human NuRD (MTA1/2/3, HDAC1/2, and RbAp46/48). Full-length, flag-tagged sponge p55BR was expressed in human epithelial kidney cells (HEK 293T) and immunoprecipitations of cell lysates with an anti-flag antibody were followed by western

blot analysis to identify the human NuRD components interacting with the flag-tagged p55BR. Figure 3.10 suggests that despite a weak band for human MTA2, full-length *Ephydatia* p55BR of MBD2 does not significantly recruit human MTA2, RbAp48, and HDAC2.

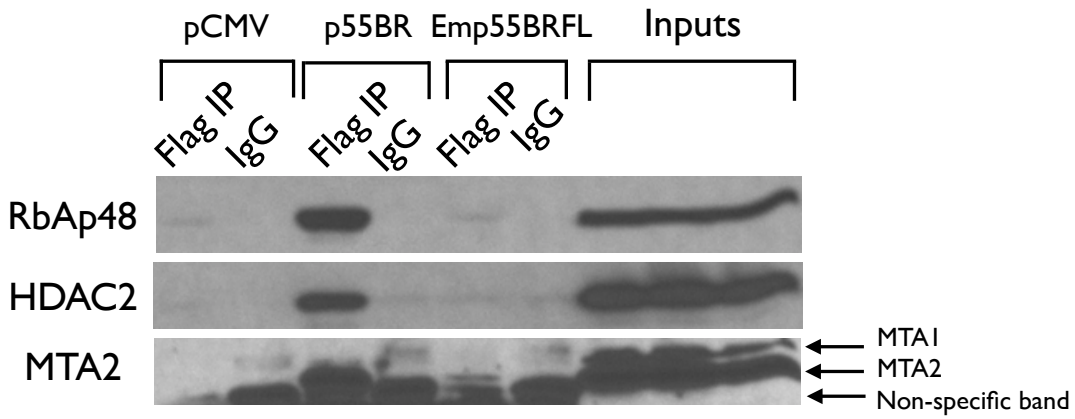


Figure 3.10: *Ephydatia* MBD2 p55BR does not recruit human NuRD components. The *Ephydatia* MBD2 p55BR-pCMVTag2B plasmid and an empty vector control were transiently transfected into high-transfection-efficiency HEK 293T cells.

3.4.7 RNAi knockdown of EmMBD2 causes atypical growth patterns in Ephydatia muelleri.

Preliminary data from the RNAi-induced knockdown of the *Ephydatia* MBD2/3 in developing sponges carried by out a collaborator resulted in an atypical growth phenotype early in development. This phenotype was characterized by abnormal growth patterns at the leading edge of tissue formation and failure to form complete structures associated with early developmental steps (i.e., canal structures and osculum).

3.5 DISCUSSION

The leap from unicellularity to multicellularity represents a major evolutionary step,

and sponges exemplify an early and successful evolutionary attempt at a collective existence. But, it is currently unknown to what extent the evolution of morphological complexity was driven by changes in gene content or regulation. Therefore, as the most basally extant organism in the animal lineage, Poriferans represent a ripe target of epigenetic investigation. The epigenetic capabilities of this most primitive and ancient multicellular organism not only represent a base set for all animals, but also provides insight into the epigenetic machinery present in the last common ancestor to all metazoans.

Here, we have presented biophysical analyses pertaining to the ability of MBD2 in the freshwater desmosponge, *Ephydatia muelleri,* to bind methylated DNA and recruit NuRD components. These studies underscore how changes in MBD2 amino acid content over time may have influenced the emergence of the NuRD complex in response to a changing DNA landscape. We specifically find that: (1) the capacity to carry out methylation-specific epigenetic mechanisms is present in sponges; (2) *E. muelleri* MBD2 and p66a can bind, albeit weakly, through their respective coiled-coil domains, despite minimal pre-formed helical content; (3) cross-species pull-downs show the *E. muelleri* MBD2 p55BR does not recruit human NuRD core components MTA2, Rbap46/48 or human HDAC1/2; (4) RNAi knockdown of MBD2 in *E. muelleri* results in abnormalities to growth and development.

Previous work has established that helical content can modulate binding affinity of the MBD2 and p66α coiled-coil domains (16). The CD and ITC data suggest the importance of pre-formed helical content regarding binding affinity and specificity. The *Ephydatia* MBD2-p66α coiled-coil interaction further supports the notion that differences

in helical content dictate high affinity coiled-coil binding for the human orthologs, and thus, overall NuRD complex formation. As we have observed, though, sponge MBD2 and p66α need not display high helical content in order to interact. Since MBD2 homologs (e.g., MBD3, MBD3L1, MBD3L2) are not found in the genomes of *Amphimedon* or *Ephydatia*, competition for p66α binding is likely reduced in the context of forming different NuRD complexes. Thus, high affinity binding may have become necessary only when other isoforms emerged.

MBD2 homologs emerge at the same time as the vertebrate methylation pattern (4), which includes largely unmethylated CGIs, suggesting a role for these proteins in regulating genes with unmethylated CGIs. Data collected from the RNAi-mediated knockdown of MBD2 supports the notion that sponges have a capacity for interpreting methyl-cytosines as an epigenetic mark. *Ephydatia* MBD2 targets methylated sites with similar high-affinity binding observed in human and chicken MBD2 isoforms. Sponges also lack homologs of MBD2 that may alter the distribution, and thus the functional consequence, of MBD2 on DNA. Therefore, sponge MBD2 may not need to bind DNA as tightly in order to bring about a physiologic response.

The role of DNA methylation and methyl-cytosine binding domains has clearly changed during the course of metazoan evolution. Expansion of the methylation landscape in vertebrate genomes created an environment in which new dynamic properties of regulatory proteins were required; expansion of the first MBD's functional role may have occurred concomitant to changes in the DNA methylome. At the dawn of animals, however, the methylation landscape may have initially dictated little need for variations of methyl-cytosine binding domain proteins. There may have been little

evolutionary pressure for the presence of different NuRD complexes that show varying degrees of DNA methylation selectivity and provide distinct functional roles.

MBD2-directed targeting to mCpGs clearly serves an important developmental role for *Ephydatia*. It is however unclear if the developmental abnormalities observed in *Ephydatia* result from actions directly mediated by an intact NuRD complex. Despite the observations collected from our cross-species experiments, it remains reasonable to expect the *Ephydatia* p55BR to recruit *Ephydatia* NuRD components in sponge cells. Differences in amino acid composition between the human and sponge proteins in question may prevent *E.m-H.s.* binding, but not necessarily between the sponge proteins. Human and *Ephydatia* RbAp share 82% amino acid identity and 80% identity is shared between the human and sponge HDAC whereas only 42% identity exists between those of MTA2. Only 30% identity is shared between the MBD2 p55BR of both species. Though sequence differences likely dictate the results observed in these studies, pull-down analysis in sponge cells is a necessary next step to determine if NuRD formation occurs in *Ephydatia muelleri*.

It is also possible that a NuRD complex may not form in its entirety in *E. muelleri*. That supposition, however, does not preclude the formation of a functional complex that targets methylated DNA and carries out chromatin remodeling. Our observations support the possibility of such a complex. Furthermore, the NuRD complex is unique in that it couples two separate DNA-related functions, whereas most co-repressor complexes perform only one action on DNA (23). It is reasonable that a sponge precursor to the NuRD complex may involve only a single regulatory mechanism. However, we cannot

conclude that MBD2 does not direct NuRD complex assembly in *E. muelleri* without an ability to perform pull-down assays in sponge cells.

The paths leading to novel protein functions likely proceed through states in which isoforms and relatively unstable structures emerge and coexist (24); one or more of these forms may perform new tasks beneficial for the organism resulting in the evolution of a new biological function. This scenario might very well be represented by the subtle changes over time in the MBD2 p55BR and the coiled-coil regions of p66α, MBD2, and the remainder of the MBD family as well. Indeed, the expansion of the MBD family appears to correlates with the evolution of their C-terminal coiled-coil regions.

Research continues to show that NuRD activity largely depends on its specific subunits and their interactions. Despite that which is known about the NuRD complex, much about the mechanisms of component recruitment, assembly, and localization to specific genomic targets remain unclear. These mechanisms undoubtedly influence subunit binding affinity and genomic selectivity by the complex, and therefore constitute an essential area of study regarding this important genetic regulator. Continued study of the evolution of NuRD subunits over time may elucidate these mechanisms. Future studies may also uncover how the changing epigenetic landscape, associated with both the metazoan-eumetazoan and the invertebrate-vertebrate transition allowed for, or was mediated by, the evolution of the NuRD complex.

3.6 REFERENCES

1. Tweedie S, Charlton J, Clark V, Bird A. Methylation of genomes and genes at the invertebrate-vertebrate boundary. *Molecular and Cellular Biology*. 1997;17:1469.
2. Bird A, Wolffe AP. Methylation-induced repression--belts, braces, and chromatin. *Cell*. 1999;99:451.
3. Sarda S, Zeng J, Hunt BG, Yi SV. The evolution of invertebrate gene body methylation. *Molecular Biology and Evolultion*. 2012;29:1907.
4. Hendrich B, Tweedie S. The methyl-CpG binding domain and the evolving role of DNA methylation in animals. *Trends in Genetics*. 2003;19:269.
5. Bird AP. Gene number, noise reduction and biological complexity. *Trends in Genetics*. 1995;11:94.
6. Coutinho CC, Fonseca RN, Mansure JJ, Borojevic R. Early steps in the evolution of multicellularity: deep structural and functional homologies among homeobox genes in sponges and higher metazoans. *Mechanisms of Development*. 2003;120:429.
7. Wainright PO, Hinkle G, et al. Monophyletic origins of the metazoa: an evolutionary link with fungi. *Science*. 1993;260:340.
8. Rivera A, Winters I, Rued A, et al. The evolution and function of the Pax/Six regulatory network in sponges. *Evolution and Development*. 2013;15:186.
9. Srivastava M et al. The Amphimedon queenslandica genome and the evolution of animal complexity. *Nature*. 2010;466:720.

10. Marhold J, Brehm A, Kramer K. The Drosophila methyl-DNA binding protein MBD2/3 interacts with the NuRD complex via p55 and MI-2. *BMC Molecular Biology*. 2004;5:20.
11. Zhang Y, Ng HH, et al. Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. . *Genes and Development*. 1999;13:1924.
12. Brackertz M, Gong Z, Leers J, Renkawitz R. p66alpha and p66beta of the Mi-2/NuRD complex mediate MBD2 and histone interaction. *Nucleic Acids Research*. 2006;34:397-406.
13. Feng Q, Cao R. Identification and functional characterization of the p66/p68 components of the MeCP1 complex. *Molecular and Cellular Biology*. 2002;22:536-46.
14. Lai AY, Wade PA. Cancer biology and NuRD: a multifaceted chromatin remodelling complex. *Nature Reviews Cancer*. 2011;11:588-96.
15. Gnanapragasam MN, Scarsdale JN, et. al. p66Alpha-MBD2 coiled-coil interaction and recruitment of Mi-2 are critical for globin gene silencing by the MBD2-NuRD complex. *Proc Natl Acad Sci U S A. 2011 May 3;108(18):7487-92. doi:*. 2011;108:787-92.
16. Walavalkar NM, Gordon N, Williams DC, Jr. Unique features of the anti-parallel, heterodimeric coiled-coil interaction between methyl-cytosine binding domain 2 (MBD2) homologues and GATA zinc finger domain containing 2A (GATAD2A/p66α). *Journal of Biological Chemistry*. 2013;288:3419-27.

17. Kingston RE, Chen CA, Rose JK. Calcium phosphate transfection. *Current Protocols in Molecular Biology*. 2003:Chapter 9:Unit 9.1.

18. Mason JM AK. Coiled coil domains: stability, specificity, and biological implications. *Chembiochem*. 2004;5:170-6.

19. Zitzewitz JA, Bilsel O, et. al. Probing the folding mechanism of a leucine zipper peptide by stopped-flow circular dichroism spectroscopy. *Biochemistry*. 1995;34:12812-9.

20. Hadley EB, Testa OD, et al. Preferred side-chain constellations at antiparallel coiled-coil interfaces. *Proceedings of the National Academy of Sciences*. 2008;105:530-535.

21. Lacroix E, Viguera AR, and Serrano L. Elucidating the folding problem of a-helices: Local motifs, long-range electrostatics, ionic strength dependence and prediction of NMR parameters. . *Journal of Molecular Biology*. 1998;284:173-191.

22. Chakrabartty A, Kortemme T, et al. Aromatic side-chain contribution to far-ultraviolet circular dichroism of helical peptides and its effect on measurement of helix propensities. *Biochemistry*. 1993;32:5560.

23. Allen HF, Wade PA, Kutateladze TG. The NuRD architecture. *Cellular and Molecular Life Science*. 2013;70:3513.

24. Gaspari Z, Nyitray L. Coiled coils as possible models of protein structure evolution. *BioMol Concepts*. 2011;2:199-210.

25. Hayes D, Philo J, Laue T. SedNTerp: Interpretation of sedimentation data version 1.X 2000 lines of Visual Basic code. 1994.

26. Schuck P. Size distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys J*. 2000;78;1606.

27. Lacroix E, Viguera AR, Serrano L. Elucidating the folding problem of alpha-helices: local motifs, long-range electrostatics, ionic-strength dependence and prediction of NMR parameters. *J. Mol. Biol*.1998;284;173.

Chapter 4- Summary

Changes to the DNA methylation landscape during the invertebrate to vertebrate transition may have accompanied evolutionary pressure for expansion of the MBD family. Global methylation patterning likely necessitates a larger ensemble of functionally diverse MBD proteins so as to expand regulatory capacity. Much like an increase in the quantity and complexity of data contained within a database requires an expansion of the database management systems that organize and analyze the data within that collection of information. Or, if one views DNA as a manuscript, alterations to the story may increase both the cast of characters and the story's complexity; more people may become involved in interpretation of the story's account of events.

As with the works of Shakespeare, we are still developing our understanding of the meaning of the stories provided to us by nature, and interpreting the menagerie of characters encoded within the script. Also like Shakespeare, nature has left us several enigmatic characters and plots we have yet to fully comprehend. The studies presented in this thesis, *Evolution And Divergence Of The Structural And Physical Properties Of DNA Binding By Methyl-Cytosine Binding Domain Family Members 2 And 3,* pertain primarily to two characters in the DNA methylation story. We more specifically provide details regarding the mystery surrounding MBD3 function and the MBD2-mediated capacity of primitive animals to carry out methylation-specific epigenetic mechanisms.

In chapter two, 'Probing the dynamic distribution of bound states for methyl-cytosine binding domains on DNA, we provide additional details regarding the DNA binding properties of MBD2 and MBD3. More specifically, this study provides details

regarding previously unidentified MBD3 binding properties and potential biological function. Studies establish the two proteins as important epigenetic regulators, and both have been analyzed extensively for possible functional differences and similarities. But although MBD2 has been well characterized, MBD3 continues to be scrutinized regarding its structure and biological function. We found that MBD3 does modify the distribution of MBD2 on DNA such that MBD2 spends less time on CpG sites. Thus MBD3 could help prevent gene silencing by MBD2 at unmethylated gene promoters and enhancers depending on the relative concentration of the two proteins and cellular context. These findings correlate with the observations that both MBD2 and MBD3 are found at unmethylated CGIs, but MBD2 binds with much greater affinity and likely excludes MBD3 from methylated CGIs. The data establishes a structural basis for the relative distribution of MBD2 and MBD3 on genomic DNA and help explain their observed occupancy at CpG-rich promoters. Further *in vivo* detail is necessary, though, to better assess MBD3 capacity for modulating MBD2 distribution at these sites and the functional consequence of such activity.

In chapter three, 'Sponge MBD2 targets methylated DNA and recruits NuRD components', we demonstrate that sponges have an MBD2-mediated capacity for binding methylated DNA sites and recruiting NuRD components *in vitro*. We further show that knockdown of MBD2 in the freshwater desmosponge, *Ephydatia muelleri*, promotes an abnormal growth phenotype, illustrating the importance of the protein for sponge development. Thus, at the dawn of animals, molecular machinery was in place allowing for methylation-specific epigenetic regulation that influences developmental processes. Further assessment should involve the ability of *E. muelleri* MBD2 to recruit sponge

NuRD components *in vitro* and *in vivo* along with analysis of NuRD complex functionality *in vivo*.

One might associate protein specificity with a lack of structural and functional versatility over evolutionary time. Indeed, proteins with absolute functional specificity and structures are likely to respond less readily to selective pressures. Structural variability consequent to mutation, gene duplication, or other processes, can lead to structural flexibility that endows proteins with functional promiscuity and/or, over time, allow structural ensembles to diverge into groups separated by distinct structures and functions. This situation may have allowed for the divergence of the MBD family from the ancestral MBD2/3. The observed flexibility in the finger-like projection connecting the central two strands of the β-sheet of MBD2 and MBD3 may exemplify a relationship between structural flexibility and evolvability or, at least, functional promiscuity. Expansion of both the MBD family and the NuRD functional repertoire may also have been promoted by subtle changes over time within the p55BR and coiled-coil regions of MBD2 and MBD3.

Much remains to be learned regarding the evolutionary and biochemical foundations of epigenetic processes. Hopefully, this dissertation provides some useful insights into the emergence and evolution of the key epigenetic regulators presented and how those molecular entities influence heritable changes in gene activity.

Vita

Jason Matthew Cramer was born on July 4, 1976, in Louisa County, Virginia, and is an American citizen. He graduated from Louisa County High School, Mineral, Virginia in 1994. He received a Bachelor of Science in Biology from Virginia Commonwealth University, Richmond, Virginia in 1999 and subsequently taught chemistry, biology, earth science, and A.P. Biology at Thomas Dale High School in Chesterfield County for nine years. During that time he earned a Master of Education in Educational Leadership with emphasis in administration and supervision from Virginia Commonwealth University in 2009.